EXHIBIT 2

EXHIBIT D

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 1

UNITED STATES DISTRICT COURT
NORTHERN DISTRICT OF CALIFORNIA
SAN FRANCISCO DIVISION

_____
Richard Kadrey, et al.,   )
Individual and Representative)
Plaintiffs,   )   CASE NO.
)   3:23-cv-03417-VC
      -against-   )
)
Meta Platforms, Inc.,   )
      Defendant.   )
_____)

*** HIGHLY CONFIDENTIAL ***
ATTORNEYS' EYES ONLY
VIDEO-RECORDED DEPOSITION OF
MELANIE KAMBADUR

Cooley, LLP
55 Hudson Yards
New York, New York 10001

09/17/2024
9:07 a.m. (EDT)

REPORTED BY: MONIQUE CABRERA

_____

DIGITAL EVIDENCE GROUP
1730 M Street, NW, Suite 812
Washington, D.C. 20036
(202) 232-0646

Page 2

1   09/17/2024
2   9:07 a.m. (EDT)
3
4   VIDEO-RECORDED DEPOSITION OF MELANIE KAMBADUR,
5 held at Cooley, LLP, 55 Hudson Yards, New York,
6 New York, before Monique Cabrera, Shorthand
7 Reporter, and Notary Public of the State of New
8 York.
9
10
11
12
13
14
15
16
17
18
19
20
21
22

Page 3

1 A P P E A R A N C E S
2 ON BEHALF OF PLAINTIFFS SARAH SILVERMAN,
3 CHRISTOPHER GOLDEN AND RICHARD KADREY:
4
5   Christopher K.L. Young, Esquire
6   Joseph Saveri Law Firm
7   601 California Street, Suite 1505
8   San Francisco, California 94108
     PHONE: 415.500.6800
9   E-MAIL: Cyoung@saverilawfirm.com
10   -AND-
11   Holden Benon, Esquire
     Joseph Saveri Law Firm
12   601 California Street, Suite 1505
     San Francisco, California 94108
13   PHONE: 415.500.6800
     E-MAIL: Hbenon@saverilawfirm.com
14   -AND-
15   Matthew Butterick, Esquire
16   Matthew Butterick Attorney at Law
     1920 Hillhurst Avenue, Suite 406
17   Los Angelos, California 90027
     PHONE: 213.422.4871
18   E-MAIL: Mb@buttericklaw.com
19 ON BEHALF OF DEFENDANT:
     Mark Weinstein, Esquire
20   Cooley LLP
     3175 Hanover Street
21   Palo Alto, CA 94304-1130
     PHONE: 650.843.5007
22   E-MAIL: Mweinstein@cooley.com

Page 4

1   A P P E A R A N C E S (CONT'D)
2   ON BEHALF OF DEFENDANT:
3   Liz Stameshkin, Esquire
4   Cooley LLP
5   3175 Hanover Street
6   Palo Alto, CA 94304-1130
7   Phone: 650.843.5121
8   E-MAIL: Lstameshkin@cooley.com
9
10   ALSO PRESENT:
11   Daniel Ortega, Legal Videographer - Digital
12   Evidence Group
13   Nikki Vo, In-house counsel for Meta
14
15
16
17
18
19
20
21
22

1 (Pages 1 to 4)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 9

1   THE VIDEOGRAPHER: We are now on the
2   record. My name is Daniel Ortega and I am
3   the legal videographer for Digital Evidence
4   Group.
5   Today's date is September 17, 2024,
6   and the time is 9:07 a.m. This video
7   deposition is being held at 55 Hudson Yards,
8   New York, New York, in the matter of Kadrey,
9   et al. versus Meta Platforms, Inc.
10   The deponent today is
11   Melanie Kambadur.
12   Counsel, please identify yourselves
13   for the record.
14   MR. YOUNG: Good morning. My name
15   is Christopher Young on behalf of plaintiffs,
16   Sarah Silverman, Christopher Golden, and
17   Richard Kadrey. With me are Holden Benon and
18   Matthew Butterick.
19   MR. WEINSTEIN: Mark Weinstein from
20   the law firm of Cooley, LLP. I'm here with
21   Liz Stameshkin, also from the Cooley law
22   firm. And Nikki Vo, who is in-house legal at

Page 10

1   Meta, for the defendant and the witness.
2   THE VIDEOGRAPHER: The court
3   reporter today is Monique Cabrera and
4   will   now swear in.
5   MELANIE KAMBADUR, called as a witness, having
6   been first duly sworn by a Notary Public of the
7   State of New York, was examined and testified as
8   follows:
9   THE WITNESS: Yes, I do.
10   COURT REPORTER: Can you state your
11   name and address for the record.
12   THE WITNESS: Melanie Kambadur. And
13   address?
14   COURT REPORTER: Yes.
15   THE WITNESS: 14 West 76th Street,
16   Apartment 2F, New York, New York 10023.
17   EXAMINATION
18   BY MR. YOUNG:
19   Q.   Good morning, Ms. Kambadur.
20   A.   Good morning.
21   Q.   So can you please state your full
22   name for the record?

Page 11

1   A.   Yes. My full name is
2   Melanie Rae Kambadur.
3   Q.   And your address was
4   ███████████████████████████████████?
5   A.   Yes.
6   Q.   Okay. And where are you presently
7   employed?
8   A.   At Meta.
9   Q.   Meta Platforms?
10   A.   Yes.
11   Q.   Okay. And what is your current
12   position at Meta?
13   A.   I'm a research engineering manager.
14   Q.   Is there a particular division you
15   work for?
16   A.   I work in the gen AI org.
17   (Reporter clarification.)
18   THE WITNESS: Gen AI.
19   BY MR. YOUNG:
20   Q.   And how long have you been in that
21   role?
22   A.   I have been a research manager at

Page 12

1   Meta for three years in -- since March.
2   Subtracting six. Six months. And in the gen AI
3   organization for about the last year and a half.
4   Q.   I'm sorry. You said you were in the
5   gen AI division for the last year and a half?
6   A.   Yes. Approximately.
7   Q.   And what division were you in before
8   you were in the gen AI division?
9   A.   In a group called FAIR.
10   Q.   And that's Facebook AI Research?
11   A.   Yes. I believe they changed it to
12   Fundamental AI Research at some point as well.
13   Q.   But when you were there, it was
14   Facebook AI Research?
15   A.   I don't recall exactly when they
16   flipped the name.
17   Q.   Okay. Have you ever had your
18   deposition taken before?
19   A.   No.
20   Q.   All right. And you may have covered
21   some of this with your counsel in preparation for
22   today. I want to -- but I want to just go over

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 13

1   some basic ground rules of deposition today.
2   If you don't understand a question,
3   please let me know and I will try to rephrase it.
4   Otherwise, I will assume you understood the
5   question; is that fair?
6   A.   Yes.
7   Q.   If you need to take a break, please
8   let me know. As long as a question is not
9   pending, I will do my best to accommodate; is
10   that fair?
11   A.   Yes.
12   Q.   And you took an oath today, right?
13   A.   Yes.
14   Q.   Okay. To tell the truth?
15   A.   Yes.
16   Q.   You understand that even though we
17   are here in a conference room taking a
18   deposition, this deposition is being conducted as
19   if you were testifying in court before a judge
20   and a jury?
21   Do you understand that?
22   A.   Yes.

Page 14

1   Q.   Is there anything we should know
2   about you that might inhibit your ability from
3   testifying today truthfully and to the best of
4   your ability?
5   A.   No.
6   Q.   So, for example, you're not on any
7   medication or any alcohol that might affect your
8   ability to answer questions here today truthfully
9   and to best of your ability?
10   A.   No, I'm not.
11   Q.   And have you ever been convicted of
12   a felony?
13   A.   No.
14   Q.   Have you ever been convicted of a
15   crime that involves an element of untruthfulness,
16   for example, fraud, embezzlement?
17   A.   No.
18   Q.   And do you understand that your
19   testimony today is being recorded by video?
20   A.   Yes.
21   Q.   And that your testimony is being
22   transcribed by a court reporter?

Page 15

1   A.   Yes.
2   Q.   Okay. Are you represented by
3   counsel here today?
4   A.   Yes.
5   Q.   Who are your counsel?
6   MR. WEINSTEIN: Mark Weinstein for
7   Cooley and the Cooley law firm.
8   BY MR. YOUNG:
9   Q.   Mr. Weinstein is representing you?
10   A.   Yes.
11   Q.   Do you understand him to be
12   representing you or Meta?
13   A.   Both.
14   Q.   And do you have any another counsel
15   representing you today?
16   A.   Yes.
17   MR. WEINSTEIN: Let record reflect
18   that the witness is gesturing towards Nikki
19   Vo, in-house legal for Meta.
20   BY MR. YOUNG:
21   Q.   Were you also gesturing to
22   Ms. Stameshkin?

Page 16

1   A.   Yes.
2   Q.   Are you paying for your counsel to
3   appear here today?
4   A.   No.
5   Q.   Did you prepare to testify today?
6   A.   What do you mean by that?
7   Q.   Did you meet with your attorneys to
8   prepare for your testimony here today?
9   MR. WEINSTEIN: You can answer the
10   question "yes" or "no."
11   A.   Yes.
12   BY MR. YOUNG:
13   Q.   How many times?
14   MR. WEINSTEIN: Again, you can
15   answer this line of questioning, just don't
16   reveal the substance of any conversations you
17   may have had with counsel.
18   A.   We had two prep meetings.
19   BY MR. YOUNG:
20   Q.   Who did you meet with, without
21   divulging the substance of your conversation with
22   counsel?

4 (Pages 13 to 16)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur

Page 17

1   A.   I met with Liz Stameshkin and Mark
2 Weinstein, Nikki Vo and a couple other folks from
3 our Meta legal counsel.
4   Q.   Were any non-attorneys present at
5 those preparation meetings?
6   A.   Not that I know of.
7   Q.   Do you remember the names of the
8 other individuals at these prep sessions?
9   A.   I remember Michelle Woodhouse. I
10 don't recall the others.
11   Q.   Michelle Woodhouse, do you
12 understand her to be legal counsel at Meta?
13   A.   Yes.
14   Q.   Did you speak about this deposition
15 with anyone but legal counsel?
16   A.   Yes. I told my boss and one or two
17 members of my team that I wouldn't be here today.
18   Q.   Did you talk about any of the
19 substance of your testimony today with anyone by
20 your counsel?
21   A.   No.
22   Q.   Who is your boss currently?

Page 18

1   A.   Manohar Baluri.
2   Q.   Can you please spell her name?
3   A.   His name is M-A-N-O-H-A-R; last name
4 B-A-L-U-R-I.
5   Q.   Who are the team members you spoke
6 with about your deposition today?
7   A.   I told Drew Mahajan.
8   Q.   Can you spell that, please?
9   A.   I can try the correct spelling, but
10 I think it's M-A-H-A-J-A-N.
11   Q.   Without going into the substance of
12 your discussions, during those prep sessions for
13 this deposition today, did you review any
14 materials?
15   MR. WEINSTEIN: You can answer the
16   question "yes" or "no."
17   A.   Yes.
18 BY MR. YOUNG:
19   Q.   Did looking at any of the materials
20 that you reviewed to prepare for your deposition
21 today -- did any of those materials refresh your
22 memory of any events that took place in the past?

Page 19

1   MR. WEINSTEIN: You can answer the
2   question "yes" or "no."
3   A.   Yes.
4 BY MR. YOUNG:
5   Q.   If yes, what about your memory was
6 refreshed in reviewing that material?
7   A.   Essentially I just recall some
8 documents that we had written.
9   Q.   Did your review of that material
10 remind you of any individuals that you have
11 conversed with in the past?
12   MR. WEINSTEIN: Object to form.
13   Vague.
14   You can answer.
15   A.   Yes.
16 BY MR. YOUNG:
17   Q.   Any specific conversations you may
18 have had with people?
19   A.   Yes.
20   Q.   Projects you may have worked on?
21   A.   I mean, yes, but it's not like -- I
22 recalled the projects beforehand.

Page 20

1   MR. YOUNG: Counsel, I am going to
2   make a request under FR 8612 for the
3   materials that refreshed the witness's
4   recollection.
5   MR. WEINSTEIN: Under the Sporck
6   case, I don't think you've established a
7   foundation for the production of those
8   materials. It's not simply enough that the
9   witness has had recollection reflected.
10   These documents have to also form the basis
11   of her testimony.
12   So under the Sporck v Peil case, the
13   Third Circuit case, I respectfully decline
14   the request on the ground that the
15   identification of documents reveals any work
16   product of counsel.
17   MR. YOUNG: Noted.
18   Counsel, we will just note -- we
19   made the counsel -- we reserve the right to
20   renew that request during the course of the
21   examination.
22   MR. WEINSTEIN: I understand.

5 (Pages 17 to 20)

Page 21

```
 1   BY MR. YOUNG:
 2        Q.   Ms. Kambadur, I do want to -- I'm
 3   sorry.  Is it Kambadur or Kambadur?
 4        A.   Kambadur.
 5        Q.   Kambadur.  I apologize for
 6   mispronouncing your name.
 7             So you hold a Ph.D.
 8   in computer science from Columbia University here
 9   in New York City, right?
10        A.   Yes.
11        Q.   You also hold an MS in computer
12   science from Columbia as well?
13        A.   Yes.
14        Q.   You hold a BS in computer science
15   from Indiana University, am I correct?
16        A.   That's correct.
17        Q.   You graduated from Indiana
18   University with highest honors?
19        A.   I don't recall.
20        Q.   It was a very impressive GPA,
21   something like 3.92, right?
22        A.   I don't recall.
```

Page 22

```
 1        Q.   Now, when did you start working at
 2   Meta?
 3        A.   I believe -- I am trying to remember
 4   if it was March of 2021 or March of 2020.
 5   March '21, I think.
 6        Q.   Before that you worked at a
 7   start-up, right?
 8        A.   Correct.
 9        Q.   It was a podcast company, podcast
10   preference company?  How would you describe it?
11        A.   Yes.  Like a podcast-related
12   start-up.
13        Q.   Before that you worked for a company
14   called Oscar Health; is that right?
15        A.   Yes.
16        Q.   What -- strike that.
17             What were your responsibilities when
18   you were working at Oscar Health, O-S-C-A-R?
19        A.   I started as a software engineer and
20   later I became a software engineering manager.
21        Q.   Did you work in natural language
22   processing?
```

Page 23

```
 1        A.   Some of the work was in natural
 2   language processing, yes.
 3        Q.   As well as in search?
 4        A.   Yes.
 5        Q.   Did you start at Meta at -- in March
 6   of 2021, does that sound about right?
 7        A.   That's sounds correct.
 8        Q.   I believe we covered this before but
 9   you started in the Facebook AI Research group,
10   right?
11        A.   Yes.
12        Q.   Do you remember who was your direct
13   report at that time?
14        A.   I had more than one direct report.
15        Q.   Do you recall how many?
16        A.   Not precisely, no.
17        Q.   Do you recall how many people were
18   in the FAIR group?
19        A.   No.  It changed over time as well.
20        Q.   Would you say more or less a dozen?
21        A.   More.
22        Q.   More or less than 20?
```

Page 24

```
 1        A.   More.
 2        Q.   When did you obtain your Ph.D. from
 3   Columbia?
 4        A.   I believe January 2016.
 5        Q.   While you were a Ph.D. candidate, do
 6   you recall the number of internships or other
 7   employment -- strike that.
 8             While you were at Ph.D. at Columbia,
 9   did you hold a number internships?
10        A.   Yes.
11        Q.   And while you were a Ph.D.
12   candidate, did you work for Google?
13        A.   Yes.
14        Q.   Intel?
15        A.   Yes.
16        Q.   Microsoft?
17        A.   Yes.
18        Q.   So you mentioned that you currently
19   work in generative AI; is that right?
20        A.   That's correct.
21        Q.   What is generative AI?
22        A.   I am not sure I have the perfect
```

Page 25

1 definition, but when an AI model produces content
2 as opposed to just classifying or producing like,
3 say, a single number.
4   Q.   I'm sorry. Did you say producing a
5 single number?
6   A.   Gen AI would produce more open-ended
7 content than other forms of AI which may be, for
8 example, classifying.
9   Q.   Would you consider large language
10 models to be a type of generative AI?
11   A.   Yes.
12   Q.   What about image diffusion models?
13   A.   Yes.
14   Q.   So when you first started at Meta,
15 you were a research manager with the FAIR group,
16 right?
17   A.   Yes.
18   Q.   What were your principal job
19 responsibilities as a research manager for the
20 FAIR group?
21   A.   I was supporting a team of research
22 engineers.

Page 27

1 by the word "weights" when you said "learned
2 numeric weights"?
3   A.   I'm not sure how to define it
4 precisely. But essentially, numeric -- a set of
5 numbers that we learn that define the model's
6 knowledge.
7   Q.   Did FAIR have a mission statement
8 when you joined it?
9   A.   That sounds familiar, yes.
10   Q.   Do you recall what that mission
11 statement was?
12   A.   No.
13   Q.   And do you have any general
14 recollection of what it was?
15   MR. WEINSTEIN: Object to form.
16   A.   Not really. I -- I would have to
17 look it up.
18 BY MR. YOUNG:
19   Q.   Okay. So at some time at FAIR, did
20 you start working on large language models?
21   A.   Yes.
22   Q.   And how long were you in FAIR when

Page 26

1   Q.   To develop large language models?
2   A.   When I first started, we were
3 working mostly on dialog models.
4   Q.   What is a dialog model?
5   A.   Models specifically tuned to produce
6 chat or dialog.
7   Q.   And when you say the word "model,"
8 do you say -- are you meaning some sort of
9 generative AI system?
10   A.   I'm not sure what you mean by
11 "generative AI system."
12   Q.   So what -- what do you mean when you
13 say "model"?
14   A.   I mean a set of learned numeric
15 weights that can take input and produce output.
16   Q.   And what is a -- so you said the
17 phrase "learned numeric weights."
18   What is a weight file?
19   MR. WEINSTEIN: Object to form.
20   A.   I'm not sure what you mean exactly.
21 BY MR. YOUNG:
22   Q.   Can you define for me what you meant

Page 28

1 you started working on large language models?
2   A.   Are -- how are we defining me
3 working on large language models?
4   Q.   So when you were first assigned to
5 work on a project that could relate to a large
6 language model?
7   A.   I'm not sure I was ever really
8 assigned, precisely.
9   Q.   So how was work divided -- how was
10 work assigned at FAIR?
11   A.   When I first joined, it was very
12 bottoms up and researchers selected the projects
13 that they worked on.
14   Q.   So at some point, did you become
15 interested in working on a project that that
16 could be useful for a large language model?
17   A.   A project -- sorry. A project that
18 could be useful for a large language model?
19   Q.   Large language --
20   A.   What do you mean by that?
21   Q.   For example, developing a large
22 language model.

7 (Pages 25 to 28)

Page 29

1   A.   I -- yes, I -- I guess I became
2 interested in it.
3   Q.   And do you remember about how long
4 you were with the FAIR group when that first
5 happened?
6   A.   I don't recall when I personally
7 became interested.
8   Q.   And what is a large language model?
9   A.   It's a model, as we discussed, which
10 has a large number of parameters, a large
11 capacity to learn and is generally text-based
12 image training data inputs and outputs.
13   Q.   So you used the word "training
14 data." What is training data?
15   A.   I mean, it's data that we use to
16 train a model.
17   Q.   And what do you mean by "training a
18 model"?
19   A.   The model sees the data, updates its
20 weights.
21   Q.   So you used that the word "weights"
22 again.

Page 30

1   So do weights affect how a model
2 behave?
3   A.   What do you mean, how a model
4 behaves?
5   Q.   So what is the purpose of weights in
6 a generative AI model?
7   A.   It's a representation to hold
8 knowledge of the model.
9   Q.   And when you say "knowledge," do you
10 mean the information contained within the model?
11   A.   What do you mean by information
12 contained within the model?
13   Q.   Do you mean the data that the model
14 was trained on when you say the word "knowledge"?
15   A.   No, not exactly. It doesn't contain
16 the data directly in those weights. It contains
17 learned relations, for example.
18   Q.   Perhaps approximation's may be a
19 better word?
20   A.   Not sure. What do you mean by
21 "approximations"?
22   Q.   Are you familiar with the phrase

Page 31

1 "lossy compression"?
2   A.   Yes.
3   Q.   Would it be fair to say that
4 generative AI models and lossy compressed
5 portions of their dataset?
6   MR. WEINSTEIN: Object to form.
7   A.   I'm not sure that's how I would
8 define it.
9 BY MR. YOUNG:
10   Q.   Have others that you've worked with
11 at Meta used the term "lossy compression" to
12 describe AI models?
13   A.   No idea. I have not specifically
14 heard it, that I recall.
15   Q.   Would you say training data is
16 important in developing an AI model?
17   A.   Yes.
18   Q.   Why?
19   A.   Training data is how the model
20 learns new information.
21   Q.   Is it important to curate the
22 training data that goes into a model?

Page 32

1   MR. WEINSTEIN: Object to form.
2 Vague.
3   A.   What do you mean by "important"?
4 BY MR. YOUNG:
5   Q.   Have you selected training data that
6 is fed into a model?
7   A.   By "selected" do you mean
8 specifically -- what -- yeah.
9   What do you mean by "selecting"?
10   Q.   So does the training data used to
11 develop a model affect how a model would perform?
12   A.   Yes.
13   Q.   So you wouldn't feed just any data
14 into a model, use my data to train a model,
15 right?
16   A.   Someone could.
17   Q.   But you wouldn't, right?
18   A.   I guess it dep- -- would depend on
19 the purpose. But generally, we do select
20 specific data, yes.
21   Q.   High quality data?
22   A.   Generally, yes.

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 33

1   Q.   What do you -- what do you
2   understand me to mean when I say "high quality
3   data"?
4   A.   Actually, I'm not 100 percent sure
5   what you mean by high quality data.
6   Q.   Is "high quality data" a term you
7   use in your work as AI developer?
8   A.   Yes, we have described data as high
9   quality levels. But I'm not sure if you have the
10   same definition as me.

Page 34

6   BY MR. YOUNG:

18   BY MR. YOUNG:
19   Q.   Would you say it's now part of your
20   responsibility -- strike that.
21   At some point, was it part of your
22   responsibility to evaluate datasets for use in

Page 35

1   training large language models?
2   A.   Not directly. I have managed people
3   who were doing the evaluation and I sometimes
4   reviewed their evaluations.
5   Q.   So by "review evaluations," do you
6   mean someone who reported to you, would evaluate
7   a dataset for use in training and then you would
8   say yes, no, this would be something we would
9   include?
10   A.   I did not solely make the yes/no
11   decisions. The first part of that statement,
12   that people who report to me would run
13   experiments and evaluate models is correct.
14   Q.   Who else -- so you said you were not
15   the person who solely made the decision to
16   include certain datasets.
17   Who else would have input into that
18   decision?

                            - like, are you asking about a
22   specific model?

Page 36

1   Q.   No. Just generally.
2   A.   Generally, members of the research
3   team   .

16   BY MR. YOUNG:
17   Q.   Are you familiar with the Llama
18   models?
19   A.   Yes.
20   Q.   What are they?
21   A.   They are a series of large language
22   models that we trained at Meta.

Page 37

1   **Q.   How many Llama models are there?**
2   A.   What do you mean exactly by that?
3   Do you mean how many models have ever been
4   trained or do you mean how many models have there
5   been released?
6   **Q.   How many models have been released?**
7   A.   We have had, to my memory, four
8   series of models, each with multiple sized
9   releases.
10   **Q.   By series of models, for example,**
11   **are you familiar with the Llama 1 series of**
12   **models?**
13   A.   Yes.
14   **Q.   Back then it was just called Llama,**
15   **right?**
16   A.   We did call that model Llama in the
17   paper.
18   **Q.   Was an early name for the Llama**
19   **model Genesis?**
20   A.   There was an internal name for Llama
21   of Genesis, yes.
22   **Q.   And the Genesis team, they were**

Page 38

1   individuals who worked with Llama mentioned --
2   what would become the Llama models?
3   A.   I can't remember if we referred to
4   them as Genesis team or not.
5   **Q.   Do you recall who was on the Genesis**
6   **team?**
7   A.   I recall some of the individuals.
8   **Q.   For example, Mr. Hugo Touvron?**
9   A.   Yes.
10   **Q.   Touvron, T-O-U-V-R-O-N.**
11   **What about Ms. Joelle Pineau?**
12   A.   I am sorry, who?
13   **Q.   Joelle Pineau?**
14   A.   She was not directly on that team.
15   **Q.   What about Luke Zettlemoyer?**
16   A.   He was not directly on that team but
17   may have been consulted by that team. I am not
18   sure.
19   **Q.   Z-E-T-T-L-E-M-O-Y-E-R.**
20   **What about Tim Dettmers?**
21   A.   I don't recall.
22   **Q.   Now, when Meta released each series**

Page 39

1   of models, does Meta usually release a research
2   paper with that model?
3   A.   Three out of the four times we
4   released a research paper so far.
5   **Q.   Does that correspond with each**
6   **series of models?**
7   A.   Roughly.
8   **Q.   So, for example, when the Llama 1**
9   **series of models was released, there was research**
10   **paper released with it?**
11   A.   Yes.
12   **Q.   Are you familiar with that research**
13   **paper?**
14   A.   Somewhat.
15   **Q.   Have you read it?**
16   A.   I believe I have read it, at least
17   parts of it.
18   **Q.   We are going to hand you what's been**
19   **premarked as Plaintiffs' 1.**
20   **(Whereupon, Kambadur Exhibit 1,**
21   **"LLaMA: Open and Efficient Foundation**
22   **Language Models," was marked for**

Page 40

1   identification.)
2   BY MR. YOUNG:
3   **Q.   It's a document bearing Bates Stamp**
4   **Meta Kadrey 00070 to 00078 beginning Bates.**
5   **Ms. Kambadur, do you recognize this**
6   **document?**
7   A.   I recognize the paper, yes.
8   **Q.   And this is a document titled**
9   **"LLaMA: Open and Efficient Foundation Language**
10   **Models"; is that correct?**
11   A.   That's correct.
12   **Q.   Was this the paper that was released**
13   **with the Llama 1 series of models?**
14   A.   It appears to be.
15   **Q.   Now, is it Meta's regular business**
16   **practice to produce research papers**
17   **contemporaneously with the model series release?**
18   MR. WEINSTEIN: Object to form.
19   You can answer.
20   A.   I am not sure I can speak more
21   broadly for the company, but as I said, we did
22   research these research papers for Llama.

10 (Pages 37 to 40)

Page 41

1   BY MR. YOUNG:
2   Q.   For Llama 1, Llama 2 and Llama 3,
3   correct, for each of those series of models?
4   A.   We released for Llama 1, Llama 2 and
5   I think we called it Llama 3.1, but I don't
6   recall what it was called internally.
7   Q.   Do you know if this paper was
8   contemporaneously released with the Llama 1 --
9   strike that.
10   Do you know if this paper was
11 released around the same time as the Llama 1
12 series model was released?
13   A.   I believe so.
14   Q.   That was some time in February of
15 2023; is that correct?
16   A.   That sounds right but I don't recall
17 the exact month.
18   Q.   Now, did you have any role in
19 drafting this paper?
20   A.   I don't believe so.
21   Q.   Did you see any drafts of this
22 paper?

Page 42

1   A.   Yes.
2   Q.   Did you see any drafts of this paper
3   that refer to -- strike that.
4   Was there an early draft of this
5   paper that did not use the word "Llama" but used
6   "Genesis" instead?
7   A.   Possibly. I don't recall.
8   Q.   Did you do any work on the Llama 1
9   series of models?
10   A.   Not directly.
11   Q.   What do you mean by "not directly"?
12   A.   For example, I supported Naman
13 Goyal, who was a co-author on this paper.
14   Q.   Do you recognize the names of all of
15 individuals listed as authors of this paper?
16   A.   I am not sure precisely what you
17 mean by that, but if they are on the list of
18 authors then, yes, this sounds correct, to my
19 memory, of the author list.
20   Q.   Do you know each of these
21 individuals?
22   A.   I, at least, know of them. I

Page 43

1   wouldn't say I know each of them well.
2   Q.   Do you know who Hugo Touvron is?
3   A.   Yes.
4   Q.   Does he work for Meta?
5   A.   Yes.
6   Q.   What is his role at Meta?
7   A.   He is a research scientist.
8   Q.   Did he have principal responsibility
9   for developing the Llama 1 series of models?
10   MR. WEINSTEIN: Object to form.
11 Vague.
12   You can answer.
13   A.   What do you mean by "principal
14 responsibility"?
15 BY MR. YOUNG:
16   Q.   Did he have managerial
17 responsibility over the development of the Llama
18 1 series of models?
19   A.   No.
20   Q.   Who would that person be?
21   A.   There were multiple managers.
22   Q.   How many?

Page 44

1   A.   I don't recall precisely.
2   Q.   Can you give me their names?
3   A.   I recall Aurelien Rodriguez and
4   Armand Joulin who are co-authors on this paper.
5   And as I said, I was also one of the managers of
6   the team.
7   Q.   Of the team, but you had no direct
8   role in development model, right?
9   A.   What do you mean by a "direct role."
10   Q.   Well, you said you didn't work in
11 the Llama 1 series of models directly.
12   Do you recall that?
13   A.   I mean that I was not implementing
14 or running experiments myself.
15   Q.   But you were managing or supervising
16 people?
17   A.   I was managing one of this team.
18   Q.   Who were you managing?
19   A.   Naman Goyal.
20   Q.   Naman Goyal reported to you?
21   A.   Yes.
22   Q.   What was Naman Goyal's role on the

11 (Pages 41 to 44)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 45

1 **Llama 1 team?**
2   A.   He was mostly helping with training
3 efficiency.
4   **Q.   What is training efficiency?**
5   A.   Making the model train on a --
6 making the model more effectively to utilize the
7 GPUs it was trained on.
8   **Q.   GPU is general processing units?**
9   A.   Yes.
10   **Q.   That's hardware?**
11   A.   Yes.
12   **Q.   I am sorry. Are they actually**
13 **graphics processing?**
14   A.   I am sorry. Yes. Correct. Thank
15 you.
16   **Q.   So Aurelien Rodriguez, he is no**
17 **longer at Meta; is that correct?**
18   A.   That is correct.
19   **Q.   Is he working for a company called**
20 **Mistral?**
21   A.   I don't think so, but I am not sure.
22   **Q.   I am sorry. Actually, is Mr. Lample**

Page 46

1 **that is actually at Mistral?**
2   A.   Yes. Guillaume, yes, is at Mistral.
3   **Q.   I ask that you forgive me for my**
4 **terrible French.**
5   **Along with Mr. Timothee Lacroix?**
6   A.   I don't recall for sure.
7   **Q.   Mr. Lample, what was his role at**
8 **Meta?**
9   A.   He was also a research scientist.
10   **Q.   And Armand Joulin, what is his role?**
11   A.   I believe research director.
12   **Q.   So research director.**
13   **Did Mr. Joulin have managerial**
14 **responsibility over others on -- listed as**
15 **authors on this paper?**
16   A.   Honestly, I forget who was on his
17 management team at the time.
18   **Q.   Who is Xavier Martinet?**
19   A.   He is another research scientist at
20 Meta.
21   **Q.   Do you know who Faisal Azhar is?**
22   A.   I believe he was in either project

Page 47

1 management or program management.
2   **Q.   Who was Edouard Grave?**
3   A.   He was another research scientist.
4   **Q.   So you can set that document aside.**
5   **Now, at some point later in 2023,**
6 **Meta released another series of models; is that**
7 **correct?**
8   A.   Yes.
9   **Q.   What would be called Llama 2 series**
10 **of models, right?**
11   A.   We released the Llama 2 series of
12 models in 2023, yes.
13   **Q.   Some time in July of 2023?**
14   A.   That sounds correct.
15   **Q.   And was there a research paper**
16 **released contemporaneously with the model -- the**
17 **Llama 2 series was released?**
18   A.   Yes.
19   **Q.   I'm going to hand you what's been**
20 **premarked as Plaintiffs' Exhibit 2. This is a**
21 **document bearing Bates Meta Kadrey 0000001.**
22   **(Whereupon, Kambadur Exhibit 2,**

Page 48

1   **"Llama 2: Open Foundation and Fine-Tuned**
2   **Chat Models", was marked for identification.)**
3 BY MR. YOUNG:
4   **Q.   Ms. Kambadur, you can just keep the**
5 **exhibits. I might be cross-referencing some of**
6 **them later today, so.**
7   **Ms. Kambadur, do you recognize this**
8 **document?**
9   A.   Yes.
10   **Q.   What is this document?**
11   A.   It appears to be our Llama 2 paper.
12   **Q.   Now, you are listed as one of the**
13 **authors of this paper, correct?**
14   A.   Yes.
15   **Q.   Along with many of the names --**
16 **well, you are on the second-to-bottom row; is**
17 **that right?**
18   A.   That's correct.
19   **Q.   Right next to Ms. Angela Fan?**
20   A.   Yes.
21   **Q.   Who is Angela Fan?**
22   A.   She's a research scientists at Meta.

12 (Pages 45 to 48)

Page 49

1   Q.   Do you know who Shruti Bhosale is?
2   A.   Yes.
3   Q.   Who is Shruti?
4   A.   She is a research engineer at Meta.
5   Q.   What about David -- David Esiobu, do
6   you know who that is?
7   A.   Yes.
8   Q.   Who is Mr. Esiobu?
9   A.   He's also a research engineer at
10   Meta.
11   Q.   What about Nikolay Bashlykov, do you
12   know who that is?
13   A.   Yes.
14   Q.   What is his role at Meta?
15   A.   A research engineer.
16   Q.   Can you please turn to page 4 of the
17   document.
18   Do you see where in the middle of
19   the page it says "We are releasing the following
20   models to general public for research and
21   commercial use"?
22   A.   Yes.

Page 50

1   Q.   What is the difference between
2   research use and commercial use?
3   A.   I think we meant it as companies who
4   are licensed to use the models, not just research
5   groups.
6   Q.   By "license" do you mean for a fee?
7   A.   No.
8   Q.   What do you mean by "license"?
9   A.   I'm not sure I know the -- the
10   precise legal definition, but we had an open,
11   i.e., free license for other companies to use it
12   under certain constraints.
13   Q.   What constraints were they?
14   A.   I don't recall the precise
15   constraints.
16   Q.   Was the Llama 1 series models ever
17   released to the public?
18   A.   Yes.
19   Q.   And by -- strike that.
20   So "constraints," that implies some
21   sort of conditions, right?
22   A.   Yes.

Page 51

1   Q.   What conditions were -- do you
2   recall any examples of those conditions that
3   companies would sign up for in order to use the
4   Llama series of models?
5   A.   I don't recall precise constraints.
6   I'd have to look at it. But I remember, for
7   example, there were constraints about not
8   causing -- not -- not using it for harm against
9   minors, I believe. Something like that in there
10   or not using it more generally. But, again, I
11   would have to look at the actual license to
12   recall.
13   Q.   So was Meta just giving the Llama 2
14   models away without expectations of anything in
15   return?
16   MR. WEINSTEIN: Object to form.
17   Vague.
18   You can answer.
19   A.   What do you mean "without
20   expectations of anything in return"?
21   BY MR. YOUNG:
22   Q.   Was Meta getting anything out of

Page 52

1   licensing the Llama 2 series of models to
2   companies?
3   A.   Our intent was to -- I don't know
4   the -- the full ramifications of licensing. But
5   one of -- at least one of the intents was to
6   share more open research and to allow others to
7   have access to similar technology. Because we
8   believe that would improve the technology faster
9   for everyone.
21   Q.   But it would be fair to say that
22   Meta received something in return for licensing

13 (Pages 49 to 52)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 53

1   the Llama 2 models, right?
2   MR. WEINSTEIN: Object to form.
3   You can answer.
4   A.   I don't -- I think it was -- my
5   understanding is that it was more abstract. It
6   wasn't like a trade for data. It was more like
7   we are trying to help improve the technology for
8   everyone.
9   BY MR. YOUNG:
10   Q.   Do you remember some of the
11   companies that Meta licensed the Llama 2 license
12   to?
13   A.   No. I wasn't really involved with
14   that process.
15   Q.   What do you understand the word
16   "commercial" to mean?
17   A.   Businesses.
18   Q.   So what do you understand the
19   term -- the phrase "commercial use" to mean?
20   A.   My understanding is that it's
21   for-profit entities and not just, like,
22   non-profit research organizations.

Page 54

1   Q.   Were any of the -- strike that.
2   So by "commercial use," when you're
3   using commercial -- when you were using
4   "commercial use" in this paper, did you mean with
5   for-profit entities?
6   MR. WEINSTEIN: Object to form.
7   Calls for speculation.
8   You can answer.
9   A.   I don't remember if I helped with
10   the sentence, but I think the intent of the
11   sentence is that nonresearch groups could use it
12   as well, such as businesses.
13   MR. YOUNG: And, Counsel, I please
14   ask you to limit your objections to form or
15   instruct the client not to -- the witness not
16   to answer.
17   MR. WEINSTEIN: Yeah, I don't think
18   it's a requirement under our rules. I'll
19   take your request under advisement.
20   MR. YOUNG: Okay. Thank you,
21   Counsel.
22   BY MR. YOUNG:

Page 55

1   Q.   Do you know who wrote that sentence?
2   A.   No.
3   Q.   Do you know who the primary author
4   of this document was?
5   A.   I don't know what you mean by
6   "primary author."
7   Q.   Do you know who wrote the majority
8   of the words that ended up in this paper?
9   A.   No, I don't know.
10   Q.   Do you know who wrote the first
11   draft of this paper?
12   A.   I'm not sure, actually.
13   Q.   Was it Mr. Touvron?
14   MR. WEINSTEIN: Object to form.
15   A.   I -- I am not sure.
16   BY MR. YOUNG:
17   Q.   So out of the names on the very
18   first page of this document, who would know what
19   you meant by the words "commercial use" in this
20   paragraph?
21   A.   I'm not sure.
22   Q.   You're not sure? Okay.

Page 56

1   Now, you licensed Llama to -- to
2   for-profit entities; is that right?
3   A.   Yes, I believe so.
4   Q.   What about Llama 1? Was Llama 1
5   licensed to for-profit entities?
6   A.   As far as I know, no.
7   Q.   Was Llama 1 largely an internal
8   model?
9   A.   What do you mean by "largely an
10   internal model"?
11   Q.   Was it generally -- would you
12   characterized it largely as for commercial --
13   research use? Excuse me.
14   A.   I'm sorry. Those are two different
15   definitions, internal versus research use.
16   Which one do you mean?
17   Q.   Well, strike that. Let me start a
18   new line of questioning.
19   So do you see the sub 1 under that
20   sentence we've been looking at?
21   A.   I'm sorry. Under which sentence?
22   Q.   That sentence, "we are releasing the

14 (Pages 53 to 56)

Page 57

1 following models to the general public for
2 research and commercial use." There are two
3 bullet -- there are two numbered bullet points.
4   Do you see that, page 4? I
5 apologize.
6   A.   I see a footnote, if that's what
7 you're -- or sorry.
8   You're talking about the bullet
9 under?
10   Q.   Yes.
11   A.   Yeah, sorry. Yes, I see that.
12   Q.   Okay. It says: One, Llama 2, an
13 updated version of Llama 1 trained on a new mix
14 of publicly available data.
15   Do you see that?
16   A.   Yes.
17   Q.   So what does it -- what does it mean
18 by Llama 2 is an updated version of Llama 1?
19   A.   It means that we added more to the
20 model to improve the model.
21   Q.   So does that mean that everything
22 that was in Llama 1 is necessarily included in

Page 58

1 Llama 2?
2   A.   No.
3   Q.   Why not?
4   A.   I don't -- I'm not sure what you
5 mean by "why not," but...
6   Q.   Well -- okay. So when you said you
7 added more to Llama 1 to make Llama 2, more what?
8   A.   For example, we further aligned the
9 model with more annotation data.
10   Q.   More training data?
11   A.   A specific type of training data,
12 yes.
13   Q.   Did you add other training data into
14 Llama 2?
15   A.   I don't recall actually.
16   Q.   So it says that Llama 2 was trained
17 on, quote, a new mix of publicly available data.
██   ████████████████████
██   ██████████████
██   ████████████████
21   Q.   Do you remember if this paper
22 describes what datasets were used to train the

Page 59

1 Llama 2 model?
2   A.   I would have to review to see.
3   Q.   Take a couple of seconds to review.
4   MR. WEINSTEIN: After the question,
5 can we take a break so my colleague could
6 leave?
7   A.   I do see at least some of the
8 datasets are reported.
9 BY MR. YOUNG:
10   Q.   Can you point me to where?
11   A.   Table 6 on page 11.
12   MR. YOUNG: Why don't we go off the
13 record real quickly so Ms. Vo can leave.
14   THE VIDEOGRAPHER: The time right
15 now is 10:06. We are off the record.
16   (Whereupon, a brief recess was
17 taken.)
18   THE VIDEOGRAPHER: Time right now is
19 10:16 a.m. We are back on record.
20 BY MR. YOUNG:
21   Q.   Welcome back, Ms. Kambadur.
22   Is there a difference between

Page 60

1 pre-training and training data?
2   A.   Pre-training is a more precise term,
3 pre-training data.
4   Q.   What is pre-training?
5   A.   Pre-training refers to the stage of
6 training where we use a large amount of
7 unsupervised data.
8   Q.   What do you mean by "unsupervised"?
9   A.   In this context, unsupervised means
10 that we have data that doesn't include labels.
11 So it will be like a piece of text versus a piece
12 of -- yes, just a piece of text essentially.
13   Q.   Now, before we left for the break,
14 we were looking at what was marked as Plaintiffs'
15 Exhibit 2, the Llama 2 model or the Llama 2
16 paper.
17   Do you remember that?
18   A.   Yes.
19   Q.   Now, if can you please go to page 5
20 of that paper, please. There's a section on
21 pre-training.
22   Do you see that?

15 (Pages 57 to 60)

Page 61

1   A.   Yes.
2   Q.   And Section 2.1 is on pre-training
3   data.
4   Do you see that?
5   A.   Yes, I do.
6   Q.   The first sentence says: "Our
7   training corpus includes a new mix of date from
8   publicly available sources, which does not
9   include data from Meta's products and services."
10   Do you see that?
11   A.   Yes.
12   Q.   What do you mean by publicly
13   available -- strike that.
14   What do you understand the term
15   "publicly available sources" to mean?
16   A.   I understand it to mean data which
17   is publicly available, for example, on the
18   Internet.
19   Q.   From any sources on the Internet?
20   A.   Are you asking like could the
21   training data be from any sources or could
22   publicly available data be from any sources?

Page 62

1   Q.   Let me rephrase that question.
2   So by "publicly available," could
3   that data, for example, just be downloaded from
4   the Internet?
5   MR. WEINSTEIN: Object to form.
6   A.   Could you download data that is
7   publicly available?
8   BY MR. YOUNG:
9   Q.   Yes.
10   A.   Yes.
11   Q.   Could that data be free?
12   A.   You are asking can data which is
13   publicly available be free?
14   Q.   Yes.
15   A.   Yes.
16   Q.   Can you describe some of the
17   publicly available data that been referred to
18   here in this paper?
19   A.   For example, we use web crawl data.
20   Q.   Are you familiar with what Common
21   Crawl is?
22   A.   Yes.

Page 63

1   Q.   What is Common Crawl?
2   A.   It is a -- it may also be an
3   institution, I am not sure, but I am familiar
4   with it as a data source that is crawled web
5   pages.
6   Q.   Is it a -- do you understand Common
7   Crawl to be a series of datasets published by an
8   organization called Common Crawl?
9   A.   I can't recall the organization's
10   name, but it may be also be called Common Crawl.
11   Q.   Do you understand generally what
12   type of information is included in the Common
13   Crawl data sources?
14   A.   I have a high level understanding.
15   Q.   What is your high level
16   understanding of what's included in there?
17   A.   That there are web documents.
18   Q.   So documents scraped from the
19   Internet?
20   A.   Yes.
21   Q.   Do you know if there any copyrighted
22   material that's been scraped and included in the

Page 64

1   Common Crawl data sources?
2   MR. WEINSTEIN: Object to form.
3   A.   I do not know.
4   BY MR. YOUNG:
5   Q.   Do you think there could be?
6   MR. WEINSTEIN: The same objection.
7   You can answer.
8   A.   It's possible.
9   BY MR. YOUNG:
10   Q.   Did you use Common Crawl to
11   pre-train the Llama models?
12   A.   Yes.
13   Q.   Do you recall how you obtained the
14   Common Crawl data sources for the purposes of
15   pre-training Llama models?
16   A.   No, I do not recall.
17   Q.   Did someone at FAIR download the
18   Common Crawl data sources for use as pre-training
19   materials for Llama models?
20   MR. WEINSTEIN: Object to form.
21   A.   It's possible.
22   BY MR. YOUNG:

16 (Pages 61 to 64)

Page 65

1   Q.   Did someone at Meta obtain the
2   Common Crawl data sources for use as pre-training
3   data for the Llama models?
4   A.   At some point we must have obtained
5   it. I am not sure exactly.
6   Q.   Did you pay for the Common Crawl --
7   did Meta pay for the Common Crawl data sources
8   for use as pre-training material?
9   A.   To my knowledge, no. It's publicly
10 available.
11   Q.   Did Meta license the Common Crawl
12 data sources for use as pre-training material for
13 the Llama models?
14   MR. WEINSTEIN: Objection to form.
15   A.   I don't know what exactly you mean
16 by "license." Can you define that?
17 BY MR. YOUNG:
18   Q.   Well, you said that the Llama 2
19 models were distributed under your license,
20 right?
21   A.   Yes.
22   Q.   Do you understand if Meta licensed

Page 66

1   with Common Crawl in order to use the Common
2   Crawl data sources for the purposes of
3   pre-training?
4   MR. WEINSTEIN: Object to form.
5   A.   I believe there might be a license
6   attached to Common Crawl.
7   BY MR. YOUNG:
8   Q.   What about the works inside the
9   Common Crawl data sources? Do you know if there
10 is license attached to those?
11   A.   It's possible for some of them.
12   Q.   Do you know if Common Crawl licensed
13 with any of the works that it scraped in order to
14 include it in the data source?
15   MR. WEINSTEIN: Object to form.
16   A.   I do not know.
17 BY MR. YOUNG:
18   Q.   Do you know what a shadow library
19 is?
20   A.   I don't recall the term.
21   Q.   Do you know what Library Genesis is?
22   A.   At a high level, yes.

Page 67

1   Q.   Is it also known as Libgen?
2   A.   Yes.
3   Q.   What do you understand that to be?
4   A.   A dataset that contains books and
5   scientific articles.

s.

18   Q.   So you mentioned that you had a high
19 level of understanding of what Libgen was. Do
20 you know specifically what type of media are
21 included in Library Genesis?
22   A.   No, I don't know specifically.

Page 68

1   Q.   But you know it contains books?
2   A.   Yes.
3   Q.   Do you know if it contains
4   copyrighted books?
5   A.   I don't know the specific books on
6   Libgen.

12 BY MR. YOUNG:
13   Q.   Do you know if there are any books
14 that are written in the last 50 years on Libgen?
15   A.   I believe so. But, again, I haven't
16 directly seen -- I don't directly recall any
17 books titles.

17 (Pages 65 to 68)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 69

Page 71

18   Q.   So, for example, if you -- you know,
19 you -- you've read books, correct?
20   A.   Yes, I've read books.

Page 70

3   Q.   What is pre-process -- processing?
4 Excuse me.
5   A.   That means transforming the data
6 with a series of steps that make it more suitable
7 for training models.
8   Q.   Does that include removing -- strike
9 that.
10   What kind of steps would be involved
11 in pre-processing?
12   A.   Are you asking specifically for
13 Llama models or more generally?
14   Q.   Yeah, for the whole Llama models.
15   A.   I don't know all the steps. B

Page 72

7   Q.   Was that information actually
8 removed for data used to pre-train the Llama
9 models?
10   A.   I don't recall.
11   Q.   Do you know -- going back to Libgen.
12   Do you know if it contained fiction
13 books?
14   A.   I believe it did.
15   Q.   How about non-fiction books?
16   A.   Yes, I believe it did.
17   Q.   Are you familiar with the Project
18 Gutenberg?
19   A.   At a very high level, yes.
20   Q.   What do you understand Project
21 Gutenberg to be?
22   A.   A -- I -- I don't know the full

18 (Pages 69 to 72)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 73

1 project, but I know that they have produced a
2 dataset of books.
3   Q.   And Project Gutenberg, it's
4 available online, right?
5   A.   As far as I know, yes.
6   Q.   Do you know if Project Gutenberg
7 contains copyrighted materials?
8   A.   I'm not certain.
9   Q.   Have you ever visited the Project
10 Gutenberg website?
11   A.   Possibly. I don't recall.
12   Q.   Do you recall seeing a disclaimer
13 saying that some of the books may be copyrighted
14 and thus may be limited for commercial uses?
15   MR. WEINSTEIN: Object to form.
16   A.   I don't recall because I don't
17 recall if I ever visited the site.
18 BY MR. YOUNG:
19   Q.   Okay. Was Project Gutenberg the
20 dataset used to pre-train any of the Llama
21 models?
22   A.   I believe so.

Page 74

1   Q.   Do you remember which models?
2   A.   I don't recall for certain.
3   Q.   Was it used to train the Llama 1
4 series of models?
5   A.   That sounds possible, but I would
6 have to look it up.
7   Q.   What about the Llama 2 series of
8 models?
9   A.   I don't recall.
10   Q.   And Llama 3, what about for those
11 series of models?
12   A.   I don't recall.
13 [REDACTED]
14 [REDACTED]
15 [REDACTED]
16 [REDACTED]
17 [REDACTED]
18   Q.   Okay. So when a model is
19 pre-trained on data, is it possible to remove
20 that data from the model's knowledge base?
21   A.   What do you mean "remove from the
22 knowledge base"?

Page 75

1   Q.   So, for example, if a model -- if
2 Llama was pre-trained on Project Gutenberg --
3 well, actually, strike that.
4   What is post-training?
5   A.   Post-training, different people have
6 different definitions, but it generally could
7 refer to any training steps done after you have
8 pre-trained a foundational model.
9   Q.   And what is a foundational model?
10   A.   Generally, a model which you intend
11 to use broadly.
12   Q.   So what do you mean by "use
13 broadly"? Can you explain that phrase a little
14 bit?
15   A.   For example, which you might
16 post-train in multiple different ways. It's like
17 a foundational base model.
18   Q.   So using layman's terms, to make the
19 model more specialized for particular functions?
20   A.   A foundational could be made
21 specialized for different functions, yes.
22   Q.   And I recall you used the phrase --

Page 76

1 the term "downstream applications" earlier.
2   Is that -- would that be --
3 post-training be something you would do to make a
4 foundational model more suitable for downstream
5 applications? Would that be fair to say?
6   A.   Yes, potentially.
7   Q.   Now, what do you mean by "downstream
8 applications"?
9   A.   I mean end uses of the model.
10   Q.   So, for example, coding?
11   A.   Coding could be something that
12 someone like a coding assistant could be a use
13 case.
14   Q.   What about writing fiction?
15   A.   What about it?
16   Q.   Could that be a downstream
17 application or a use case?
18   A.   Could someone use a -- an LLM to
19 writes fiction, is what you're asking?
20   Q.   Yes.
21   A.   Potentially as an aid.
22   Q.   What about poetry?

19 (Pages 73 to 76)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 77

1   A.   Yes, models can output poems.
2   **Q.   What about legal briefs?**
3   A.   I've never tried using it to output
4 a legal brief. I imagine someone could generate
5 a document that looks similar to a legal brief.
6 But most of the language models would probably
7 not be accurate or comprehensive for an entire
8 legal brief.
9   **Q.   Because of the risk of the model**
10 **hallucinating?**
11   A.   Yes. In part, the risk of the model
12 hallucinating.
13   **Q.   And what do you understand the word**
14 **"hallucinating" to mean?**
15   A.   My understanding is that it means
16 the models is saying something -- it is producing
17 an untrue output.
18   **Q.   Do you know who Stella Biderman is?**
19   A.   Yes.
20   **Q.   Who is Stella Biderman?**
21   A.   I know that they work at Eleuther.
22   **Q.   EleutherAI, right?**

Page 78

1   A.   EleutherAI, yes.
2   **Q.   Have you met Ms. Biderman?**
3   A.   I was once on a panel.
4   **Q.   And it's a Zoom panel?**
5   A.   I don't recall the platform, but it
6 was virtual.
7   **Q.   Who is EleutherAI -- or excuse me.**
8   **What is a EleutherAI?**
9   A.   I don't know that much about the
10 company.
11   **Q.   Okay. Do you know if they released**
12 **any datasets?**
13   A.   To my recollection, yes.
14   **Q.   Is one of those datasets the Pile?**
15   A.   That sounds right, but I'm not sure.
16   **Q.   When did you first learn about**
17 **EleutherAI?**
18   A.   I don't recall.
19   **Q.   Was it before -- before or after**
20 **Llama 1's release?**
21   A.   I don't recall.
22   **Q.   Do you know who Leo Gao is?**

Page 79

1   A.   The name doesn't sound familiar.
2   **Q.   Do you know if the Pile -- well, do**
3 **you understand what generally comprises the Pile?**
4   A.   I don't recall the datasets in the
5 Pile.
6   **Q.   Do you recall a dataset referred to**
7 **as Books3?**
8   A.   I recall the dataset, yes.
9   **Q.   And do you understand Books3 to be a**
10 **dataset included in the Pile?**
11   A.   That sounds correct, but I don't
12 remember for certain.
13   **Q.   Do you remember if Books3 was used**
14 **as pre-training data for the Llama 1 series of**
15 **models?**
16   A.   Yes, I believe it was.

Page 80

17   **Q.   Do you know what is in Books3?**
18   A.   I know there are books. I don't
19 know the specific contents.
20   **Q.   I suppose the name dataset gives it**
21 **away a little bit. Do you know if it's fiction?**
22   A.   Yes, it contains fiction.

Page 81

1   Q.   Do you know if it contains
2   copyrighted works?
3   A.   I believe so.
4   Q.   What about non-fiction?
5   A.   I am not sure.
6   Q.   Do you know if Meta ever paid for
7   any of the books included in the Books3 data set?
8   MR. WEINSTEIN: Object to form.
9   A.   No, I don't know.
10 BY MR. YOUNG:
11   Q.   Do you know how many books are in
12 Books3?
13   A.   I don't recall.
14   Q.   Do you know if Meta maintains a list
15 of all the books included in Books3?
16   A.   Of all the books included, I am not
17 sure I have seen a list of titles but I don't
18 recall if that was comprehensive.
19   Q.   I apologize. I interrupted you.
20   A.   Sorry.

Page 82

4   Q.   Do you know if that document still
5   exists at Meta now?
6   A.   I am sure it does.
7   Q.   Would that document be a true and
8   accurate list of the titles used to train any of
9   the Llama models?
10   MR. WEINSTEIN: Object to form.
11   A.   I am not certain. The best source
12 of truth would be the training data files
13 themselves.
14 BY MR. YOUNG:
15   Q.   The training data files would be
16 voluminous, right?
17   A.   Would be what? I'm sorry.
18   Q.   Voluminous.
19   A.   I am sorry. I don't know what you
20 mean by that.
21   Q.   The training dataset files are quite
22 large, right?

Page 83

1   A.   Relative to what?
2   Q.   Relative to data?
3   A.   Sorry. I don't know what you mean
4   by that.
5   Q.   I apologize. I am not a data
6   person. I will try to rephrase.
7   Have you ever examined the training
8   data?
9   A.   I have seen small subsets of it.
10   Q.   Have you ever seen the entire --
11 well, do you know what is a training corpus or
12 pre-training corpus?
13   A.   It's a -- corpus is just a term that
14 describes the set of data.
15   Q.   So it would include all of the
16 pre-training datasets, would that be fair?
17   A.   That is usually what people mean
18 when they say "training corpus," yes.
19   Q.   So when I refer to -- do you
20 understand by dataset, that may be a subset of a
21 training corpus, right?
22   A.   Yes. A dataset could be a subset.

Page 84

1   Q.   So a training set comprises -- would
2   it be fair to say when you look at a data --
3   well, strike that.
4   Do you know what a JSON file is?
5   A.   Yes.
6   Q.   JSON is a file extension, right?
7   A.   Yes.
8   Q.   Would it be fair to say that some of
9   the training datasets would exist in JSON format?
10   A.   At some point it's possible that the
11 training data was in a JSON format.
12   Q.   So when you open a JSON file, it's
13 generally a bunch of text; would that be fair to
14 say?
15   A.   If -- yes, if you open it with a
16 document, you can see text on a JSON file.
17   Q.   Do you recall how many entries would
18 be in the Books3 pre-training set?
19   A.   I don't recall.
20   Q.   Thousands?
21   A.   I don't recall, and I don't know
22 also what we mean by an "entry" in this case.

21 (Pages 81 to 84)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 85

1   Q.   So you have examined -- you've seen
2 a list of -- I believe you testified earlier
3 there was list of files that are used to train
4 the Llama models.
5   Do you remember that testimony
6 earlier generally?
7   A.   I have seen a list of at least some
8 of the books used to train Llama models, yes.
9   Q.   How many entries are on that list?
10   A.   I don't recall.
11   Q.   To your best recollection, was it
12 more than 1,000?
13   A.   I don't recall but it's possible.
14   Q.   More than hundreds, though, right?
15   A.   Yes, I think so.
16   Q.   Do you know who acquired the Pile
17 for use as pre-training material for the Llama
18 models?
19   A.   I do not.
20   Q.   Do you know when you first heard of
21 the Pile being used at Meta for pre-training
22 purposes?

Page 86

1   A.   I don't remember.
2   Q.   2022?
3   A.   Perhaps earlier. I don't remember
4 which one.
5   Q.   What about Books3 specifically? Do
6 you recall when you first learned Books3 was
7 being researched at Meta for use as pre-training
8 data?
9   A.   I don't recall.
10   Q.   Was it before 2022?
11   A.   I don't recall.

19   A.   By -- actually, I don't recall
20 specifically the person.
21   Q.   Do you know if Meta maintains a list
22 of all the pre-training material used for Llama

Page 87

1 models?
2   A.   We maintain, like, configuration
3 files that point to a list of datasets and we
4 maintain the raw data.
5   Q.   So these lists of datasets, are
6 those the pre-processed datasets or the processed
7 datasets that were used for pre-training? I
8 apologize.
9   A.   Are you asking does the config file
10 point to pre- or post-process data?
11   Q.   Yes.
12   A.   Our config file would point to
13 post-process data.
14   Q.   So if one would look at the config
15 file, you would be able to identify the works
16 that would be included as pre-training data for
17 the Llama models; would that be fair to say?
18   A.   What do you mean "the works?"
19   Q.   For example, if a particular book
20 were used from Books3 where you use the
21 pre-training model, you would be able to look at
22 the config file to identify that book; would that

Page 88

1 fair?
2   A.   The config file would only give you
3 high level file names   but that
4 wasn't the precise name, just an example.
5   Q.   I understand. Thank you for
6 clarifying.
7   Do you know who Meta's competitors
8 are in terms of large language models?
9   A.   What do you mean by "competitors?"
10   Q.   For example, would OpenAI be a
11 competitor?
12   A.   I don't understand what you mean by
13 the term "competitor."
14   Q.   Do you know who OpenAI is?
15   A.   Yes.
16   Q.   Who are they?
17   A.   An AI company.
18   Q.   And did they generate an LLMs large
19 language product?
20   A.   Yes. They have a large language
21 based product.
22   Q.   And that's ChatGPT?

22 (Pages 85 to 88)

Page 89

1   A. I don't know exactly how they bundle
2 their models as products, but that's one of the
3 names of their models or products, yes.
4   Q. What about Anthropic, do you know
5 who they are?
6   A. Yes, I know of Anthropic.
7   Q. They also distribute or publish a
8 large language model known at Claude; is that
9 fair?
10   A. What do you mean by "distribute" or
11 "publish"?
12   Q. Make available for use.
13   A. Yes, they make a model called Claude
14 available for use.
15   Q. Would you consider the Llama models
16 to be competing products at ChatGPT and Claude?
17   A. What do you mean by "competing"?
18   Q. Does Meta measure the performance of
19 the Llama models compared with the other large
20 language models that are being published by, for
21 example, OpenAI and Anthropic?
22   A. Yes, we do compare performance

Page 90

1 metrics.
2   Q. Why do you compare the metrics?
3   A. We're trying to understand the
4 relative capabilities of the models.
5   Q. Why are you trying to understand the
6 relative capabilities of the models?
7   A. Because we want to have models with
8 relatively strong capabilities.
9   Q. You want your models to be better
10 than OpenAI's models, for example?
11   A. We are -- yeah. We trying to get
12 strong metrics relative to other models.
13   Q. And that's, for example, users?
14   A. I'm sorry. What do you mean?
15   Q. People who use the large language
16 models, would you want more users -- would you
17 prefer more users of Llama than, for example,
18 ChatGPT?
19   A. So the metrics I was referring to
20 earlier are more like academic benchmark numbers.
21   Q. Do you know what datasets are used
22 in ChatGPT -- excuse me. Let me strike that.

Page 91

1   Do you know what datasets were used
2 to pre-train ChatGPT?
3   MR. WEINSTEIN: Object to form.
4   A. No, I do not.
5 BY MR. YOUNG:
6   Q. Have you ever told anyone who works
7 for OpenAI the datasets that are used to
8 pre-train the Llama models?
9   A. No. I may have discussed with an --
10 I may have had discussions with an -- a former
11 employee who ended up going to OpenAI, but I
12 don't recall specifically.
13   Q. What about Anthropic, have you ever
14 discussed with anybody at Anthropic any of the
15 datasets used to pre-train any of the Llama
16 models?
17   A. Not that I recall.
18   Q. Would you consider the datasets that
19 are used to pre-train the Llama models to be
20 trade secrets?
21   MR. WEINSTEIN: Object to form.
22   A. What do you mean by "trade secrets"?

Page 92

1 BY MR. YOUNG:
2   Q. Would you ever publicly disclose the
3 datasets that were used to pre-train any of the
4 Llama models?
5   A. For Llama 2 onwards we've decided
6 not to disclose the pre-training datasets.
7   Q. Would you be subject to internal
8 discipline if you were to publicly disclose any
9 of the datasets used to pre-train any of the
10 Llama models subsequent to Llama 2?
11   MR. WEINSTEIN: Object to form.
12   A. I'm not sure, honestly.
13 BY MR. YOUNG:
14   Q. Do you recall that the Llama 1 paper
15 disclosed the pre-training datasets used for the
16 Llama 1 models?
17   A. I'd have review the paper.
18   Q. We'll do that in a second.
19   Who was that former employee of
20 OpenAI that you discussed pre-training datasets
21 with?
22   A. I said it's possible. I'm not

23 (Pages 89 to 92)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 93

1   certain I did discuss.
2   Q.   Who was that person?
3   A.   I'd have to look up the name. I --
4   I don't know.
5   Q.   Do you remember about when that was?
6   A.   I don't -- I don't recall any
7   specific conversation with this person. I'm just
8   saying it's -- it's possible he may have heard
9   about our pre-training datasets.
10   Q.   So in your role right now, you have
11 people who report to you.
12   Would that be fair to say?
13   A.   Yes.
14   Q.   Do those individuals have knowledge
15 of the datasets used to pre-train any of the
16 Llama models?
17   A.   Some of them do.
18   Q.   Now, if one of those individuals
19 were to go to, for example, OpenAI and use
20 discuss the model -- what datasets were used to
21 pre-train Llama 3, would that person be subject
22 to discipline?

Page 94

1   MR. WEINSTEIN: Object to form.
2   A.   I haven't read any of our, like,
3   employment termination documentation, so I don't
4   know.
5   BY MR. YOUNG:
6   Q.   But you have an understanding of
7   Meta's employment policies, right?
8   A.   Not in that level of detail.
9   Q.   What about data -- data sharing
10 policies? Do you have any understanding of
11 Meta's data sharing policies?
12   MR. WEINSTEIN: Object to form.
13   A.   Not a detailed understanding.
14 BY MR. YOUNG:
15   Q.   Would you consider Meta's -- the
16 datasets used to pre-train Meta's Llama models to
17 be proprietary information?
18   A.   What do you mean by "proprietary"?
19   Q.   Would you consider it to be -- do
20 you have an understanding what the word
21 "proprietary" means?
22   A.   I'm not sure what it means in a

Page 95

1   legal context.
2   Q.   What about in a general layperson's
3   context?
4   A.   It can -- maybe, like, knowledge
5   that belongs to a specific company.
6   Q.   That shouldn't be divulged. Would
7   that fair to say?
8   A.   I suppose.
9   Q.   Would the datasets used to pre-train
10 any of the Llama 2 models and subsequent models,
11 would that be information that would be free for
12 you to disclose?
13   A.   We consider it more confidential.
14 We --
15
16
17   MR. YOUNG: Counsel, I please ask
18 you to refrain from nodding while the witness
19 is answering.
20   MR. WEINSTEIN: Was that directed to
21 me?
22   MR. YOUNG: Yes.

Page 96

1   MR. WEINSTEIN: I don't think I've
2   been doing that, sir.
3   MR. YOUNG: Okay. I will note on
4   the record that you were just nodding at that
5   last response. And that was not the first
6   time I've noticed it, but I'm going to put it
7   on -- put it on the record now.
8   MR. WEINSTEIN: You can put -- you
9   can put on the record whatever you want, sir.
10   MR. YOUNG: Thank you, Counsel.
11   So we're going mark Plaintiffs' 3, a
12 document bearing Bates label Meta
13 Kadrey 000034501.
14   (Whereupon, Kambadur Exhibit 3, Team
15 off-site agenda Bates Number Meta Kadrey
16 000034501, was marked for identification.)
17 BY MR. YOUNG:
18   Q.   Do you recognize this document?
19   A.   Yes.
20   Q.   What is it?
21   A.   It looks like an agenda for a team
22 offsite.

24 (Pages 93 to 96)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 97

1   Q.   Okay. By team offsite that means --
2   well, is this -- by "team offsite," that means an
3   offsite, for example, meeting?
4   A.   It's just a colloquial term we use
5   when people from multiple offices might be
6   getting together. In this case, it was actually
7   at our Menlo Park -- or actually, was this at one
8   at New York? I forget.
9   It was at one of our Meta offices,
10   but it was some people from different offices
11   meeting together.
12   Q.   And do you remember what the general
13   subject matter of this offsite was?
14   A.   I can see from this agenda that we
15   were talking about different projects.
16   Q.   Including the Llama 2 series models?
17   A.   I don't actually recall when this
18   was, but it looks like there was at least some
19   content on Llama 2, based on the agenda.
20   Q.   Now, do you see your name on here?
21   A.   Yes.
22   Q.   And you were giving a presentation

Page 98

1   at this offsite, right, according to this agenda?
2   A.   Yes.
3   Q.   Okay. Do you recall giving a
4   presentation some time on October 24th in 2023?
5   A.   I don't recall the specific date,
6   but it sounds possible.
7   Q.   Okay. Do you recall having an
8   offsite some time around October 24, 2023?
9   A.   That sounds right.
10   Q.   Okay. Do you -- you recall
11   attending one and giving a presentation there,
12   right?
13   A.   Yes. I -- I don't remember the
14   specific dates, but based on this agenda, that
15   seems correct.
16   Q.   Now, were agendas for offsite
17   meetings created in the regular of course of
18   business at Meta?
19   A.   Were they what in the --
20   Q.   Created in the regular course of
21   business for Meta.
22   A.   I don't know exactly what you mean

Page 99

1   by that.
2   Q.   So whenever you had an offsite, was
3   there generally an -- an agenda that accompanied
4   that offsite?
5   A.   Yes. Any meeting that we have with
6   a large number of people we generally create an
7   agenda.
8   Q.   And that included off-sites that
9   were generally for business purposes, as opposed
10   to, for example leisure outings?
11   A.   You're asking if we create agendas
12   for business purpose meetings? Sorry.
13   Q.   Let me -- let me try to rephrase
14   that.
15   So looking at this agenda, this
16   agenda seems to be about business-related
17   activities, right?
18   A.   There were both business and social
19   activities, to my recollection.
20   Q.   Okay. But there were
21   business-related activities, right?
22   A.   Yes.

Page 100

1   Q.   Now, do you recall if there was a
2   PowerPoint presentation that you used when you
3   were presenting at this event?
4   A.   I don't recall for certain, but it's
5   possible there was a Google Slides presentation.
6   Q.   Okay. And Google Slides is
7   something that's like PowerPoint?
8   A.   Similar.
9   Q.   Okay. Now, whenever there was a
10   presentation at these offsites, were they
11   typically accompanied by a presentation, whether
12   it's Google Slides or some PowerPoint-like
13   analog?
14   A.   It depends how much time we set
15   aside to prepare what the topic is.
16   Q.   Okay.
17   MR. YOUNG: I'm going mark as
18   Plaintiffs' 4, a document bearing Bates label
19   Meta Kadrey 00054739.
20   (Whereupon, Kambadur Exhibit 4,
21   Presentation Bates Number Meta Kadrey
22   00054739, was marked for identification.)

25 (Pages 97 to 100)

Page 101

1 BY MR. YOUNG:
2   Q.   And before we -- take your time
3 but -- Ms. Kambadur, can you go back to
4 Plaintiffs' 3, please? It's the agenda we were
5 just looking at.
6   A.   Yes.
7   Q.   Do you see at the very top there is
8 a -- what appears to be a URL?
9   A.   Yes.
10   Q.   Do you recognize what that is?
11   A.   This hyperlink internal at
12 FD.com/wavefinder?
13   Q.   Yes.
14   A.   It looks like probably a room
15 location.
16   Q.   A room location, you mean like a
17 physical room?
18   A.   Yes, like it's a -- this MPK number
19 is the numbering and lettering system used --
20 well, I think so. I think it is a room number.
21   Q.   All right. Thank you.
22   Returning to Plaintiffs' 4, do you

Page 102

1 recognize this document?
2   A.   It looks -- yes.
3   Q.   Okay. What is it?
4   A.   I think it is the presentation we
5 gave at this off-site.
6   Q.   On or about October 23rd of 2024?
7   A.   Based on the Bates, on the first
8 slide, I think so.
9   Q.   Okay. Excuse me, 2023, right?
10   A.   Thank you.
11   Q.   I apologize.
12   A.   You are confusing me there.
13   Q.   Yes. It's funny how we can even
14 misread it when it's on the page.
15   If you look at first page, it
16 says -- second slide, it says "Agenda"?
17   A.   Yes.
18   Q.   You see your name, right?
19   A.   I see my name, yes.
20   Q.   Melanie under LLM, that's you?
21   A.   Yes.
22   Q.   That suggests that you were

Page 103

1 presenting on large language models this day,
2 right?
3   A.   It -- the one -- the main person
4 presenting the talk.
5   Q.   If you turn to the next slide?
6   A.   Yes.
7   Q.   It says "Llama 3 LLM data."
8   Do you see that?
9   A.   Yes, I see that.
10   Q.   Did that suggest to you that the
11 subject matter of your conversation that day or
12 presentation that day was the Llama 3 model?
13   A.   Yes.
14   Q.   And the data that was used in it?
15   A.   We talked a bit about the content of
16 the dataset.
17   Q.   That data set?
18   A.   Data corpus.
19   Q.   Pre-training corpus, right?
20   A.   I would have to review the talk, but
21 I recall that it was mostly on pre-training.
22   Q.   Now, can you please turn to the

Page 104

1 fifth slide. This is the one with the
2 multicolored windows. It's Llama versus the
3 competition.
4   Do you see that?
5   A.   Yes.
6   Q.   Do you understand what the slide
7 meant by "competition"?
8   A.   I think we likely meant other models
9 where we were trying to get stronger or similar
10 academic benchmarks to.
11   Q.   What are those academic benchmarks?
12   A.   The benchmarks we have used have
13 changed a lot over time but, for example, MMLU.
14   Q.   I am sorry, what?
15   A.   MMLU.
16   Q.   What is that?
17   A.   I don't recall what it stands for,
18 first the MM. Language understanding, I believe,
19 is the LU.
20   Q.   So looking at the slide here, each
21 of these windows has a word or couple words and
22 some numbers.

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 105

1   **Do you know what Anise is?**
2   A.   It is a pre-training version of
3   Llama.
4   **Q.   Does it correspond with the Llama**
5   **model that was released?**
6   A.   I believe this corresponds with the
7   Llama 1 model.
8   **Q.   What about Cinnamon?**
9   A.   There was a name for the
10   pre-training -- for another pre-training version.
11   **Q.   Another pre-training version of the**
12   **Llama model?**
13   A.   Of a Llama model, yes.
14   **Q.   Do you know which model that was?**
15   A.   I believe Llama 2.

Page 106

4   **Q.   What is Mistral?**
5   A.   I think we were referring to one of
6   their main language models at the time from
7   Mistral, the company.
8   **Q.   Mistral, the company, do you**
9   **understand that former Meta employees went on to**
10   **found Mistral, the company?**
11   A.   Yes. At least one of the founders
12   was from Meta.
13   **Q.   Do you recall who that was?**
14   A.   Guillaume Lample.
15   **Q.   GPT-4, what is that?**
16   A.   That's referring to a model from
17   OpenAI.
18   **Q.   Now, under GPT-4 at the very bottom**
19   **green window it says: Text plus image.**
20   **What does that mean?**
21   A.   That means we assumed that their
22   training data include text and images.

Page 107

1   **Q.   The pre-training data, right?**
2   A.   I believe that's what we were
3   referring to here.
16   **Q.   Generally, are more tokens better in**
17   **an LLM?**
18   A.   What do you mean by "better"?
19   **Q.   Does that make the performance of**
20   **the large language model better?**
21   A.   It depends. Generally we saw that
22   more training tokens can improve the performance

Page 108

1   of a model, yes.
2   **Q.   Do you remember presenting on this**
3   **slide?**
4   A.   Do I remember what?
5   **Q.   Presenting on this slide?**
6   A.   I don't know. I don't know what you
7   mean. Do I remember what -- the day that I
8   presented or what do you mean?
9   **Q.   Do you remember talking about this**
10   **slide during this off-site?**
11   A.   Vaguely.
12   **Q.   Vaguely.**

Page 109

2   Q.   Let me clean up that question. I
3 apologize. Strike that. Let me try a new
4 question.
5   Does this slide suggest at this
6 point in 2023 -- October 24th of 2023 -- let me
7 back up.
8   At this point Llama 3 had not been
9 released yet, October 24th of 2023?
10   A.   Correct.
11   Q.   At this point the publicly available
12 Llama version was Llama 2?
13   A.   Yes.
14   Q.   That corresponded with Cinnamon,
15 right?
16   A.   Yes.
17   Q.   At this time was the GPT-4 model
18 publicly available?
19   A.   I believe that you could access the
20 GPT-4 model through an API, yes.
21   Q.   What about the Mistral model that
22 you were referencing here?

Page 110

1   A.   It was available to access, yes.
19   Q.   At the very bottom there is a note?
20   A.   Yes.
21   Q.   This says, "token counts includes
22 epochs," right?

Page 111

1   A.   Yes.
2   Q.   What does the word "epoch" mean?
3   A.   Epoch means that we go over a piece
4 of data more than once in training.
5   Q.   So what do you mean by -- by
6 "training," do you mean pre-training?
7   A.   Yes. Pre-training.
8   Q.   So what does it mean to go over a
9 piece of data in pre-training?
10   A.   It means while we are training the
11 model and updating its weights, we show it --
12 those epoched datasets potentially more than
13 once.
14   Q.   Why would you show the model -- what
15 do you mean by show the model of the dataset?
16   A.   The model -- I'm trying to think how
17 to explain this. Hang on.
18   It means we input those tokens into
19 the, like, training algorithm into the
20 transformer network and the model does a series
21 of linear algebra steps to -- to figure out how
22 to updates its weights.

Page 112

1   Q.   So I think you said a lot there that
2 I think were terms of art. So I want to break
3 that down a little bit just to help us understand
4 a little bit.
5   So when you input tokens for
6 pre-training, is that able to work and -- well,
7 strike that.
8   What is tokenization?
9   A.   Tokenization broadly refers to
10 subdividing text into small unit -- units, which
11 can be mapped to an embedding space.
12   Q.   Is this what's referred to as the
13 neural network?
14   A.   No.
15   Q.   What is the neural network?
16   A.   The neural network probably refers
17 to the actual model.
18   Q.   Now, is tokenization, is that
19 part -- part of the post-process -- processing
20 the data before you pre-train a model?
21   A.   I'm not actually sure what we're
22 referring to as a pre-processing versus

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 113

1   post-processing, but it is part of the
2   processing, for certain.
3   Q.   So just to clarify again.
4   Post-processing happens after a model is
5   pre-trained?
6   A.   I think you're mixing that up with
7   post-training.
8   Q.   So what -- so what's the difference
9   between pre-processing and post-processing,
10 again?
11   A.   That's a good question. I think
12 it -- the term is a little bit ambiguous and
13 people sometimes use a bit interchangeably. But
14 like maybe before or after some events, but it --
15 it's sometimes used it a bit ambiguously.
16   Q.   Okay. Can you turn to the next
17 slide, please, to Slide 7. I don't if your this
18 PowerPoint is numbered.
19   A.   It's not numbered.
20   Q.   All right. Can you turn to the
21 slide that says "Goals" and it has four bullet
22 points. More data, higher qualify, more diverse

Page 114

1   and compliance. Let me know when you're there.
2   (Reporter clarification.)
3   MR. YOUNG: Sorry.
4   BY MR. YOUNG:
5   Q.   More data, higher quality, more
6   diverse and compliance, just for reference.
7   Let me know when you're there.
8   A.   Yes, I see it.
9   Q.   Okay. So do you see the first
10 bullet, it says "More data"?
11   A.   Yes.
15   Do you see that?
16   A.   Yes.
17   Q.   Okay. Do you remember presenting on
18 that bullet point?
19   A.   I have a vague recollection, just
20 like all of these slides.
21   Q.   Okay. Do you remember what that
22 bullet point meant -- strike that.

Page 115

1   Do you have an understanding of what
2   that bullet point means?
3   A.   I mean, I have an understanding of
4   what I think it means today.
5   Q.   Okay. And can you tell us the
6   understanding?
7   A.   It means that we were aiming to
8   collect additional unique tokens for Llama 3.

Page 116

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 117

Page 119

1   MR. WEINSTEIN: Exclude any
2   communications you may or may not have had
3   with counsel.
4   (Reporter clarification.)
5   MR. WEINSTEIN: Exclude any
6   communications you may or may not have had
7   with counsel.

18   MR. WEINSTEIN: Again, same -- same
19   instruction. His question's limiting to
20   non-attorneys, but if you are -- in answering
21   the question, if -- you shouldn't repeat
22   anything a lawyer may have told you, even

Page 118

Page 120

1   between a lawyer within Meta.
2   A.   It's possible we discussed. I -- I
3   don't recall.
4   MS. STAMESHKIN: I think we've been
5   going for a little over an hour.
6   MR. YOUNG: I mean, I think I can --
7   we got a couple more questions. If you can
8   give me a couple minutes.
9   MS. STAMESHKIN: Okay.
10   BY MR. YOUNG:
11   Q.   Can you turn to Slide 9, please.
12   MR. WEINSTEIN: Can you identify
13   that, please?
14   MR. YOUNG: Slide 9, this is the
15   slide that says: Additional data (raw
16   counts).
17   MR. WEINSTEIN: Got it, okay.
18   BY MR. YOUNG:
19   Q.   Let me know when you're there.
20   A.   Yes, I'm there.
21   Q.   Do you recognize this slide?
22   A.   Yes.

30 (Pages 117 to 120)

Page 121

1   Q.   What is it?
2   A.   I believe it's showing additional
3   data token counts that we were considering after
4   Llama 2.
5   Q.   To process and include in Llama 3,
6   right?
7   A.   Yes.
8   Q.   Okay. So the first row in the
9   yellow, it says: Corpus as of Llama 2.
10   Do you see that?
11   A.   Yes.
12   Q.   Okay. Can you tell me what that row
13   means?
14   A.   I think that is the size of the
15   Llama 2 corpus and tokens.
16   Q.   And that includes the pre-training
17   corpus of Llama 2, right? That represents all
18   the tokens in Llama 2, right?
19   A.   Yes, in this case I believe it's
20   referring to pre-training dataset sizes.
21   Q.   Okay. And then these rows in white,
22   right, these -- do you have any an understanding

Page 122

1   what those rows are?
2   A.   My understanding is that they were
3   additional tokens that we had collected or
4   processed.

Page 123

3   Q.   Do you recall what datasets were
4   considered in general -- in -- for inclusion in
5   this table?
6   A.   The specific datasets I would have
7   to reference our, yeah, documentation. I don't
8   recall all the datasets.
9   MR. YOUNG: I think this is a good
10   time for a break.
11   THE VIDEOGRAPHER: The time right
12   now is 11:31. We are off the record.
13   (Whereupon, a luncheon recess was
14   taken.)
15   THE VIDEOGRAPHER: The time right
16   now is 12:22 p.m. We are back on record.
17   BY MR. YOUNG:
18   Q.   Welcome back, Ms. Kambadur.
19   A.   Thanks.
20   Q.   Can you tell me what a context
21   window is?
22   A.   A context window is kind of like the

Page 124

1   size of an input that a model can understand at
2   one time.
3   Q.   Can it also relate to the length of
4   an output?
5   A.   It can relate to the length of an
6   output.
7   Q.   So when you say "input," do you mean
8   something like a prompt that a user of a large
9   language model may feed into a large language
10   model?
11   A.   The input could be a prompt from a
12   user, yes.
13   Q.   Could it be in the form of a text
14   query, for example, a question?
15   A.   Yes.
16   Q.   And the output is what the large
17   language model gives out in response to the
18   prompt, right?
19   A.   Yes. The output is what the model
20   produces.
21   Q.   And is the length of an output
22   limited by some of the technical specifications

31 (Pages 121 to 124)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 125

1 of an LLM?
2   A.   Yes. The length of an output is
3 limited.
4   Q.   And a model with a longer context
5 window would be able to, from what I understand,
6 one, understand longer prompts; would that be
7 fair to say?
8   A.   Yeah. A longer context window
9 allows you to have -- yes, an understanding of a
10 longer input.
11   Q.   A longer context window would mean
12 that LLMs would also be able to generate longer
13 outputs, right?
14   A.   Not necessarily.
15   Q.   Why not necessarily?
16   A.   There is inference time parameters
17 that affect the output length and also a model
18 may just learned patterns to typically output
19 short lengths of text or other outputs.
20   Q.   Okay. Do you recall earlier today
21 we talked about the Llama 3.1 research paper?
22   A.   Yes.

Page 126

1   Q.   Did you have a role in drafting that
2 research paper?
3   A.   A minor role.
4   Q.   Did you read the Llama 3.1 research
5 paper?
6   A.   The final version, not in entirety.
7   Q.   Do you know if the Llama 3.1
8 research paper is available online from Meta's
9 website?
10   A.   I believe it is, yes.
11   MR. YOUNG: Okay. Mark as
12 Plaintiffs' 5 a document we pulled from
13 Meta's website yesterday.
14   (Whereupon, Kambadur Exhibit 5, "The
15 Llama 3 Herd of Models," was marked for
16 identification.)
17 BY MR. YOUNG:
18   Q.   Ms. Kambadur, do you recognize this
19 document?
20   A.   Yes.
21   Q.   What is it?
22   A.   It appears to be the Llama 3 paper.

Page 127

1   Q.   It's titled "The Llama 3 Herd of
2 Models," correct?
3   A.   Yes.
4   Q.   It was published on or about
5 July 23rd of 2024?
6   A.   That sounds right.
7   Q.   And if you go to page 72 you see a
8 page bearing the header "Contributors and
9 Acknowledgements."
10   A.   Yes, I do.
11   Q.   Okay. The very first paragraph
12 defines core contributors as people who worked on
13 Llama 3 for at least two-thirds of the run time
14 on the project.
15   Do you see that?
16   A.   Yes, I do.
17   Q.   And do you see your name -- I know
18 it's a long list under the list of core
19 contributors -- somewhere in the middle of the
20 block of text, I believe the name of the authors
21 are listed in alphabetic order by first name.
22   A.   Yes, I see it.

Page 128

1   Q.   What does mean that you worked on
2 Llama 3 for at least two-thirds the run time on
3 the project?
4   A.   That means it was my full-time focus
5 for at least two-thirds of the time from when we
6 started developing Llama 3 actively.
7   MR. YOUNG: And I am sorry. Can you
8 go briefly off record. Is there a way we can
9 put a sign that says --
10   THE VIDEOGRAPHER: We are still on
11 the record. The time right now is 12:28 p.m.
12   We are off the record.
13   (Whereupon, a brief recess was
14 taken.)
15   THE VIDEOGRAPHER: Time right now is
16 12:31 p.m.
17   We are back on record.
18 BY MR. YOUNG:
19   Q.   So, Ms. Kambadur, what is the run
20 time referring to?
21   A.   Where?
22   Q.   In the page 72 that we were looking

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 129

1   at says "core contributors worked on Llama 3 for
2   at least two-thirds of the run time on the
3   project."
4   What does the run time on the
5   project mean?
6   A.   I am not precisely sure, actually.
7   Q.   Okay. Can you turn back to the
8   first page, please. Let me know when you are
9   there.
10   A.   Yes.
11   Q.   Do you see the bullet point towards
12   the bottom, it says "data"?
13   A.   Yes.
14   Q.   Okay. It says: Compared to prior
15   versions of Llama, we improved both quantity and
16   quality of the data we used for pre-training and
17   post-training.
18   Do you see that?
19   A.   Yes.
20   Q.   There's a citation to Touvron, et
21   al., 2023 A and B.
22   Do you see that?

Page 131

17   quality data is generally more valuable for
18   pre-training LLMs?
19   MR. WEINSTEIN: Object to form.
20   A.   What do you mean by "valuable"?
21   BY MR. YOUNG:
22   Q.   You will want to train an LLMs on

Page 130

1   A.   Yes.
2   Q.   Do you understand those citations to
3   be referring back to the Llama 1 and Llama 2
4   papers respectively?
5   A.   I assume, yes, but I have not
6   checked.
7   Q.   Okay. So what do you understand
8   this to mean when it says "improve both the
9   quantity and quality of the data used for
10   pre-training and post-training"?
11   A.   I am not quite sure what
12   interpretation you want other than the literal
13   one, which is that we improved the quality and
14   quantity.

Page 132

1   high quality data as opposed low quality data.
2   Would that be fair to say?
3   A.   If you had a sufficient amount of
4   high quality data, yes, you would prefer that
5   your data is more token efficient.
6   Q.   What do you mean by "token
7   efficient"?
8   A.   Meaning that you're improving the
9   performance of your metrics more effectively for
10   the same amount of data processed.

19   Q.   Can you please turn to page 4 of the
20   Llama 3 paper, please.
21   Do you see a section labelled "3
22   Pre-Training"?

33 (Pages 129 to 132)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 133

```
 1      A.   Yes.
 2      Q.   There's a section 3.1.1 web data
 3  curation.
 4           Do you see that?
 5      A.   Yes, I see that.
 6      Q.   Do you see a section that says "PII
 7  and safety filtering"?
 8      A.   Yes.
 9      Q.   Do you see in that first sentence,
10  it says:  Along with other mitigations, we
11  implement filters designed to remove data from
12  websites that are likely to contain unsafe
13  content or high volumes of PII?
14      A.   Yes.
15      Q.   What does mean to filter as used in
16  this sentence?
17      A.   It means that we do not use that
18  data for training.
19      Q.   By "training," you mean
20  pre-training, correct?
21      A.   In this case, yes.
22      Q.   And how do you filter out content
```

Page 134

```
 1  for pre-training?
 2      A.   I wasn't directly involved in the
 3  implementation, so I have only high level
 4  understanding.
 5
 6
 7
 8
 9
10  Actually, I don't -- I don't want to speak to the
11  details because I don't know exactly how we
12  implemented in this case.
13      Q.   It's fair to say there is some
14  methodology employed to remove materials used to
15  pre-train the Llama 3 model because it was either
16  unsafe or contained PII, correct?
17      A.   What do you mean by "remove"?
18      Q.   Filter.  It says:  Filters designed
19  to remove data from websites are likely to
20  contain unsafe content or high volumes of PII.
21
22
```

Page 135

```
 1
 2
 3
 4
 5
 6
 7
 8
 9
10
11
12
13
14
15
16
17
18
19      Q.   And you understand PII to mean
20  personally ident- -- personally identifiable
21  information or identifying -- not personally --
22  excuse me.
```

Page 136

```
 1      A.   Personally identifiable information
 2  is what I understand it to mean for PII.
 3      Q.   Can you turn to the next page,
 4  please.
 5           Okay.  Do you see a section called
 6  "De-Duplication"?
 7      A.   Yes.
 8      Q.   We were discussion de-duplication
 9  earlier today.
10           Do you recall that?
11      A.   Yes, I recall that.
12      Q.   Okay.  So this section describes
13  generally how the pre-training dataset or the
14  pre-training corpus was de-duplicated for
15  Llama 3; is that correct?
16      A.   Yes, that's correct.
17      Q.   Can you describe for me what URL
18  level de-duplication is?
19      A.   That means, if we saw a URL
20  associated with a document repeated more than
21  once, we would attempt to use only one version of
22  that document.
```

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 137

1   Q.   What about document level
2 de-duplication?
3   A.   That means, if we have a text
4 document that is extremely similar to another
5 text document, we attempt to use only one version
6 in our initial corpus.
7   Q.   You -- and de-duplication is
8 important because seeing the same text multiple
9 times could affect the weights of the model.
10   Would that be fair to say?
11   A.   I don't fully understand the
12 question. Any -- any text can affect the weight
13 of the models that's trained -- that the model is
14 trained on. So the question doesn't totally make
15 sense.
16   Q.   So why did you de-duplicate the
17 pre-training dataset?
18   A.   We wanted to have a set of unique
19 tokens or unique -- near unique -- sorry --
20 unique -- let me correct that -- unique documents
21 as near as we could in our initial corpus.
22   Q.   And what does "line level

Page 138

1 de-duplication" mean?
2   Do you see that in the document?
3   A.   Yes.
4   Q.   Can you tell me what that is?
5   A.   Give me a second to review this.
6   Q.   No problem. Take your time.
7   A.   Yeah. So it means that we, in some
8 cases, try to eliminate -- I'm not sure precisely
9 how we defined a line. If that -- I don't recall
10 if that was, like, a sentence or some other
11 parsing or some other chunk of text.
12   But on some chunk of text, we --
13 some span of text, we attempted to de-duplicate
14 with other similar spans of texts that were
15 extremely repeated in our training corpus.
16   Q.   So line of text, I think, appeared
17 more than six times in each bucket of 30 -- 30Ms,
18 that's 30 million documents, right?
19   A.   That's what the paper says.
20   Q.   Okay. Thank you.
21   So do you see the next section says
22 "Heuristic filtering"?

Page 139

1   A.   Yes.
2   Q.   And it says: We develop heuristics
3 to remove additional low quality documents,
4 outliers and documents with excessive
5 repetitions.
6   Do you see that?
7   A.   Yes, I see that.
8   Q.   Can you tell me what you understand
9 a low quality document to mean?
10   A.   In this case, documents indicated by
11 heuristics that may be token inefficient.
12   Q.   Does that also -- also capture
13 undesirable documents you want in your
14 pre-training set?
15   A.   It could.
16   Q.   For example, in that second bullet
17 point underneath "Heuristics filtering," it
18 discusses adult websites.
19   Do you see that?
20   A.   Yes, I see that.
21   Q.   Is that because you didn't want
22 adult content coming out of the Llama 3 model?

Page 140

1   A.   I don't recall exactly why we
2 filtered this, but adult content -- content is
3 somewhat subjective and generally, we found these
4 documents to be token inefficient anyways.
5   Q.   So why were they token inefficient?
6   A.   I don't investigate the pages
7 myself, so I can't really answer.
8   Q.   Who -- who would be able to give me
9 that answer?
10   A.   I'm not sure anyone was manually
11 reviewing the adult content documents.
19   Q.   Were there certain types of
20 documents that were considered high quality or
21 low quality?
22   A.   What do you mean by "types of

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 141

1   documents"?

[redacted]

16   Q.   So, for example, would novels be
17   considered high quality, generally?
18   MR. WEINSTEIN: Object to form.
19   A.   I -- I don't recall the specific
20   experiments we've done that we ever precisely
21   subdivided novels as a category.
22   BY MR. YOUNG:

Page 142

[redacted]

6   Q.   And when you say "categorization,"
7   what are the other types of categories you were
8   using for experiment?
9   A.   Do you mean categories we were using
10   for -- for what?
11   Q.   For pre-training.
12   So, for example, if you recall that
13   PowerPoint, we can use that to refresh your
14   recollection. That slide that included -- do you
15   remember that slide that included [redacted]
[redacted]
17   A.   Yes.
18   Q.   Those white rows?
19   A.   Yes.
[redacted]

Page 143

1   Q.   Can you please turn back to the
2   Llama 3 paper, please.
3   Could you please go to page 24 of
4   that document. Let me know when you're there.
5   A.   Yes, I'm there.
6   Q.   And you see the section that's
7   bearing header 4.3.4 "Long Context"?
8   A.   Yes, I do.
9   Q.   Okay. And this section is
10   describing the relative length of the context
11   window based on -- from Llama 3 to Llama 2; is
12   that correct?
13   A.   It says that we extent the context
14   length of Llama 3 from 8K tokens to 128K tokens.
15   And yes, I think 8K is referring to Llama 2.
16   Q.   So extending the context window of
17   Llama 3 from 8K to 128K tokens, is that a
18   significant increase, in your opinion?
19   MR. WEINSTEIN: Object to form.
20   A.   Well, I -- I'm not sure. What do
21   mean by "significant"?
22   BY MR. YOUNG:

Page 144

1   Q.   It's much longer, the context
2   window, from Llama 3 to Llama 2?
3   A.   It is longer.
4   Q.   Would you be able to describe for us
5   how much longer that context window is, 8K to
6   128K?
7   A.   128 divided by 8. I can do the
8   math.
9   Q.   Would that mean the context window
10   is 16 times longer?
11   A.   Exactly 16 -- give me a second.
12   Q.   This isn't a test. So --
13   A.   Yes. 16. 16 times longer.
14   Q.   Would that mean -- would that mean
15   that Llama 3 understand 16 times longer inputs
16   than Llama 2?
17   A.   In select cases it can understand up
18   to 128K context length.
19   Q.   What about for output? What does
20   this imply for the output of Llama 3 as compared
21   to Llama 2?
22   A.   It implies very little.

36 (Pages 141 to 144)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 145

1   Q.   Okay. You can set that document
2 aside.
3   I am going to mark as Plaintiffs' 6
4 a document bearing Bates label Meta Kadrey
5 00034707.
6   (Whereupon, Kambadur Exhibit 6,
7   Document Bates Number Meta Kadrey 00034707,
8   was marked for identification.)
9 BY MR. YOUNG:
10   Q.   Do recognize this document?
11   A.   No.
12   Q.   Then I am going to pass around the
13 metadata associated with this document.
14   MR. WEINSTEIN: I just want to be
15   the record's clear. You are not going to
16   mark it?
17   MR. YOUNG: Mark the next exhibit as
18   6A.
19   (Whereupon, Kambadur Exhibit 6A,
20   Document, was marked for identification.)
21 BY MR. YOUNG:
22   Q.   Ms. Kambadur, do you see the

Page 146

1 Document 6A?
2   A.   Yes, I do.
3   Q.   Okay. Do you see the field, "all
4 custodians"?
5   A.   Yes, I see it.
6   Q.   Do you see your name there?
7   A.   Yes.
8   Q.   So that means that this document was
9 pulled from your custodial files.
10   Do you understand that?
11   A.   I don't understand exactly what that
12 means.
13   Q.   That means that this document was
14 collected from your files during the document
15 collection process.
16   A.   So potentially on a Google Drive I
17 have access to or something like that?
18   Q.   I think you would have to talk your
19 counsel about that.
20   A.   Okay.
21   Q.   But that does not refresh your
22 recollection about whether or not you recall this

Page 147

1 document?
2   A.   I am shared a lot of documents in my
3 job.
4   Q.   You can set that document aside.
5   Does Meta have an internal
6 collaboration platform that it uses amongst
7 employees?
8   A.   Yes.
9   Q.   Would that include an internal
10 messaging application?
11   A.   Yes, we have an internal messaging
12 application.
13   Q.   Is that Workplace?
14   A.   The message application is Workplace
15 Chat.
16   Q.   You use Workplace -- I am sorry.
17 Workplace Chat -- is that?
18   A.   Chat.
19   Q.   Chat?
20   A.   C-H-A-T.
21   Q.   Do you use Workplace Chat to conduct
22 business at Meta?

Page 148

1   A.   Yes, we do.
2   Q.   And are the messages you send on
3 Workplace Chat saved by Meta?
4   A.   I believe so.
5   Q.   Do you know what memorization is?
6   A.   In what context?
7   Q.   In the generative AI context.
8   A.   Yes.
9   Q.   What -- describe for me what
10 memorization means in the generative AI context.
11   A.   Memorization means that we may have
12 the ability to output content that is similar to
13 input.
14   Q.   Would you -- memorization -- can
15 memorization -- excuse me. Let me start that
16 over.
17   Could memorization refer to the
18 phenomenon of training data appearing verbatim in
19 the generation of output?
20   A.   I am sorry. Can you repeat the
21 question?
22   Q.   Yeah. Could memorization refer to

37 (Pages 145 to 148)

Page 149

1   the repetition of training data in the output?
2   MR. WEINSTEIN: Objection to form.
3   A.   By "repetition of training data in
4   the output" --
5   BY MR. YOUNG:
6   Q.   Let me strike that.
7   Could "memorization" refer to
8   training data or pre-training data appearing
9   verbatim as output?
10   A.   Yes, it could refer to that.

22   Q.   So would it be fair to say for an

Page 150

1   order -- let me strike that.
2   Would it be fair to say that if an
3   output matches verbatim certain length with a
4   certain percentage match, that that would be
5   problematic for Meta?
6   MR. WEINSTEIN: Object to form.
7   A.   Yes. What do you mean by
8   "problematic"?
9   BY MR. YOUNG:

18   Q.   Was the output of verbatim match --
19   excuse me. Strike that.
20   Was the output of verbatim -- strike
21   that again.
22   Was the output of pre-training data

Page 151

1   verbatim as output something you tried to avoid
2   when designing your element?
3   A.   Is the output of pre-training --
4   sorry. I couldn't parse that completely.

12   MR. YOUNG: Okay. We are going to
13   mark Plaintiffs' 7 bearing Meta Kadrey
14   00054282.
15   (Whereupon, Kambadur Exhibit 7,
16   Workplace Chat document Bates Number Meta
17   Kadrey 00054282, was marked for
18   identification.)
19   BY MR. YOUNG:
20   Q.   Do you recognize this document?
21   A.   Not particularly but it looks like
22   Workplace Chat.

Page 152

1   Q.   Does this document appear to be a
2   Workplace Chat between yourself and others taking
3   place on or around February 16th of 2023?
4   A.   Yes, it does.
5   Q.   Now, Armand Joulin, do you remember
6   him, we talked about him earlier?
7   A.   Yes.
8   Q.   Can you remind me what his position
9   was at Meta at this time in 2023?
10   A.   I believe he was a research
11   director.
12   Q.   Was he above you or below you in the
13   org chart at this time?
14   A.   He was not in my direct reporting
15   line, but he had a more senior level than me.
16   Q.   These are chats that you exchanged
17   in the ordinary course of your business, right?
18   A.   I assume. I have not reviewed.
19   Q.   Just let me know when you are done.
20   A.   Okay. I reviewed the document.
21   Q.   Okay. Ms. Kambadur, does this
22   appear to be a Workplace Chat that you engaged in

38 (Pages 149 to 152)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 153

1 **with others on or about February 16th of 2023?**
2   A.   Yes.
3   **Q.   Along with Armand Joulin, right?**
4   A.   Armand is in the chat, yes.
5   **Q.   And February 16th of 2023, that was**
6 **before the Llama 1 model was released, correct?**
7   A.   I don't recall the exact release
8 dates.
9   **Q.   That's all right. So if you look at**
10 **the top message Mr. Joulin says to you -- bearing**
11 **the timestamp 2:54 PST.**
12   **Do you see that message?**
13   A.   Yes, I see that message.
14   **Q.   He says:**
17   **Do you see that?**
18   A.   Yes.
19   **Q.   What did you understand Armand to**
20 **mean when he sent that message?**
21   A.   Essentially that you want your LLM
22 to have correct association of facts that is true

Page 154

1 to your pre-training corpus.
2   **Q.   And hallucination, we discussed that**
3 **briefly, that generally means that the LLM**
4 **outputs facts that are untrue, right?**
5   A.   Yes.
6   **Q.   So Armand goes on. He says:**
10   **Do you see that?**
11   A.   Yes.
12   **Q.   Did you share Armand's concern with**
13 **using data that contained copyrighted**
14 **information?**
15   A.   I'm not even sure exactly what his
16 concern is from reviewing this.
17   **Q.   Okay. Do you think his concern with**
18 **using information that contained copyrighted**
19 **information is related to memorization?**
20   MR. WEINSTEIN: Object to form.
21   A.   It's possible but it's hard to tell
22 for sure.

Page 155

1 BY MR. YOUNG:
2   **Q.   Do you think it could be perhaps**
3 **memorization means that -- would mean that the**
4 **model may regurgitate copyrighted information?**
5   A.   I wouldn't want to speculate.
6 There's not enough context here.
7   **Q.   So Armand goes on, right? He**
8 **says -- well, before we go -- so he refers to**
9 **Common Crawl, right?**
10   **Do you see that?**
11   A.   Yes, I see the reference to Common
12 Crawl.
13   **Q.   Okay. And those -- do you**
14 **understand to mean the Common Crawl datasets we**
15 **were discussing earlier today?**
16   A.   Yes.
17   **Q.   Okay. And these are the Common**
18 **Crawl datasets that were used to pre-train likely**
19 **the Llama 1 model, given the timing of this --**
20 **this message thread, right?**
21   A.   I'm not sure which model line he's
22 referring to in this message. But it's actually

Page 156

1 unlikely to be Llama 1.
2   **Q.   Okay.**
3   A.   But it's possible -- actually, I'm
4 not -- I don't want to make a hard stance on
5 that. I'm not totally sure --
6   **Q.   Okay.**
7   A.   -- if he's -- it's referring to
8 Llama 1 or Llama 2 or some other model.
9   **Q.   But he is referring to some Llama**
10 **model, right?**
11   A.   I am not even 100 percent sure if
12 it's to a Llama model.
13   **Q.   Okay. So "e.g." -- do you see that**
14 **e.g.?**
15   A.   Yes.
16   **Q.   Do you understand what e.g. means?**
17   A.   I forget what it stands for, but
18 essentially, he's meaning it to say "for
19 example."
20   **Q.   Right.**
21   **Do you understand e.g. means -- is**
22 **the Latin abbreviation for "for example," right?**

39 (Pages 153 to 156)

Page 157

1   A.   Yes, that's what it means.
2   Q.   Okay. So do you understand Armand
3   to be expressing a problem with using Common
4   Crawl because Common Crawl contains copyrighted
5   information?
6   MR. WEINSTEIN: Object to form.
7   A.   No, that's not really my -- I don't
8   really understand that in reading this.
9   BY MR. YOUNG:
10   Q.   Okay. So Armand goes on, right? So
11   he says -- do you see where he says -- there's a
12   little caret and he says "For (1)"?
13   A.   Yes.
14   Q.   Okay. And he says: ███████████
     ███████████████████████████████
     ██████████████████████████
17   Do you see that?
18   (Reporter clarification.)
     ██████████████████████
20   BY MR. YOUNG:
21   Q.   Do you see that?
22   A.   Yes, I see that.

Page 158

1   Q.   Okay. Do you understand Armand to
2   be proposing a solution to the potential problem
3   he identifies earlier?
4   A.   What do you mean by that?
5   Q.   You respond to Armand; is that
6   right?
7   A.   Yes.
8   Q.   Okay. So the very bottom, you say:
     █████████████████████████████████
     ████████████████████████████
11   Do you see that?
12   A.   Yes, I see that.
13   Q.   Okay. And then you correct
14   yourself -- I believe you correct the typo,
15   right, right underneath of the --
16   A.   Looks like it.
17   Q.   Yep. So what did you mean by --
18   when you said that statement?
19   A.   I don't recall precisely. But I --
20   from the context, I assume it has to do with
21   filtering out private information about people.
22   Q.   Okay. Your message was about public

Page 159

1   figures from Wikipedia and -- scratch that.
2   Strike that.
3   And you were -- so you were
4   suggesting filtering out pre-training data time;
5   is that correct? Would that be fair to say?
6   A.   Yes, and I think specific to PII.
7   Q.   Okay. Would there be any reason
8   that couldn't be used for other types of
9   information at pre-training?
10   A.   Can we filter any other information
11   besides PII at pre-training?
12   Q.   Yes.
13   A.   Yes, we can sub-select what data
14   goes into a model.
15   Q.   For example, we were discussing
16   adult websites earlier, right?
17   A.   Yes, we were discussing adult
18   websites earlier.
19   Q.   Right. And how Meta wanted to
20   filter out adult content out from the
21   pre-training for Llama 3.1, right?
22   A.   I don't think that's precisely what

Page 160

1   we discussed or what was said. ██████████
     ████████████████████████████████
     ████████████████
4   Q.   And if we go further down the
5   thread, do you see your message 7:43:35 PST?
6   A.   Yes.
7   Q.   Okay. So you say: █████████████
     █████████████████████████████████
     ████████████████████████████████████
     ███████████████████████████████
12   Do you see that message?
13   A.   Yes, I do.
14   Q.   All right. "FTing," does that mean
15   fine tuning?
16   A.   Yes.
17   Q.   What is fine tuning?
18   A.   Fine tuning is a type of
19   post-training where we are trying to align the
20   model to human preferences with supervised data.
21   Q.   What do you mean by "supervised
22   data"?

Page 161

1   A.   Data where you have a -- a label on
2 the data indicating that it's a -- like,
3 indicating something about the data example.
4   **Q.   And then Armand Joulin provides a**
5 **response, right?**
6   A.   Yes.
7   **Q.   It says: "You mean directly in the**
8 **pre-training set?"**
9   **Right?**
10   A.   Yes.
11   **Q.   And then you respond: "Yes or at**
12 **generation time."**
13   A.   Looks like I did, yes.
14   **Q.   Okay. Can you explain to us what**
15 **you were advocating for when you said "Yes or a**
16 **generation type"? Can you explain the context of**
17 **that statement?**
18   A.   From this -- I mean, I don't recall
19 exactly what I meant at the time. But from this
20 context, it looks like I'm saying that ███████

Page 162

1 ████████████████████████████
████████████████████████████████
████████████████████████████████████
5 I do not mean at pre-training time.
6   **Q.   But when you say "yes," that means**
7 **at pre-training type, right?**
8   A.   Yes, we could -- yes, I -- I see.
████████████████████████████████
████████████████████████████████
████████████████████████████████
████████████████████████████████
████████████████████████████
15   **Q.   And then Mr. Lample, he responds to**
16 **you, right?**
17   A.   Yes. Guillaume Lample responded.
████████████████████████████████
████████████████████████████████
████████████████████████████████
████████████████████████████
22   **Q.   Do you agree with his assessment?**

Page 163

1   A.   It's hard to say because I don't --
2 I am not sure I understand exactly what we were
3 even proposing filtering.
4   **Q.   Okay. Can you go back up to**
5 **David Esiobu's message at 6:23:38 PST.**
6   **Do you see that?**
7   A.   Yes, I see that.
████████████████████████████████
████████████████████████████████
████████████████████████████████
████████████████████████████████
████████████████████████████████
████████████████████████████████
████████████████████████████████
████████████████████████████
████████████████████████████████
████████████████████████████████
████████████████████████████████

Page 164

████████████████████████████████
████████████████████████████████
████████████████████████████████████
████████████████████████████
████████████████████████████████
████████████████████
████████
████████████████████
██████████
10   **Q.   Or your understanding.**
11   A.   I mean, it's like -- it's a very
12 short sentence without a lot of context. So I
13 don't know if it changes my understanding.
14   (Reporter clarification.)
15   A.   It's a very short sentence without a
16 lot of context. So I -- I don't that I would say
17 it changes my understanding.
18 BY MR. YOUNG:
19   **Q.   Would, for example, if the Llama**
20 **model were to regurgitate sections of**
21 **pre-training data verbatim, would that be**
22 **undesirable memorization?**

41 (Pages 161 to 164)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 165

1   A.   Sections of pre-training data. Can
2 you define that more precisely?
3   Q.   Yes. Sure.

10   MR. YOUNG: Okay. I'm finished with
11 this document. I think we've been going on
12 for about an hour. So I think this might be
13 a good time for a short break.
14   MR. WEINSTEIN: Great.
15   THE VIDEOGRAPHER: The time right
16 now is 1:22 p.m. We are off the record.
17   (Whereupon, a brief recess was
18 taken.)
19   THE VIDEOGRAPHER: Time right now is
20 1:35 p.m.
21   We are back on record.
22   MR. YOUNG: So we passed around

Page 166

1 what's been marked as Plaintiffs' Exhibit 8.
2 It's document bearing Bates label Meta Kadrey
3 00054284.
4   (Whereupon, Kambadur Exhibit 8,
5 Workplace Chat Bates Number Meta Kadrey
6 00052484, was marked for identification.)
7 BY MR. YOUNG:
8   Q.   Ms. Kambadur, have you had a chance
9 to review that document?
10   A.   I have just skimmed it so far.
11   Q.   Do you recognize that document?
12   A.   I don't specifically recall this
13 conversation, but it looks like a discussion with
14 me and some colleagues.
15   Q.   Does this appear to be a Workplace
16 Chat between yourself, Mr. Steven Roller, and Mr.
17 Ryan Mark on or about December 2nd of 2022?
18   A.   Yes, it does.
19   Q.   Were these messages exchanged during
20 the ordinary course of business?
21   A.   Yes.
22   Q.   Okay. Do you see at the very

Page 167

1 beginning you share a document in the thread?
2   A.   I see what looks like a PNG file.
3   Q.   What is a PNG file?
4   A.   I don't know what it stands for but
5 it's typically an image.
6   Q.   The image or a screenshot, right?
7   A.   A screenshot as a type of image,
8 yes.
9   Q.   So who is Ryan Mark?
10   A.   He was a TPM that we used to work
11 with.
12   Q.   What is a TPM?
13   A.   Technical product manager.
14   Q.   What were Mr. Mark's
15 responsibilities at this time?
16   A.   I am not sure I can speak
17 comprehensively to his responsibilities because I
18 do not manage him, but he was assisting our team
19 with various operational efforts.
20   Q.   What do you mean by operations?
21   A.   For example, helping us get computer
22 resources.

Page 168

1   Q.   What about acquiring datasets?
2   A.   It looks like he was helping with
3 that from the context of these conversations, but
4 I don't recall specifically.
5   Q.   Okay. And who is Steven Roller?
6   A.   Steven Roller was a research
7 engineer on my team at the time.
8   Q.   Is Mr. Roller still with Meta?
9   A.   No.
10   Q.   Do you know where he has gone to?
11   A.   He is currently at Google DeepMind.
12 Let's just say Google, because I am exactly not
13 sure what department but Google.
14   Q.   DeepMind is Google's AI division,
15 right?
16   A.   Yes. But on a second thought I am
17 not precisely sure how they are organized
18 currently or what department he is in.
19   Q.   But do you understand at some point
20 Google DeepMind was Google's AI division?
21   A.   One of their AI divisions.
22   Q.   Thank you.

42 (Pages 165 to 168)

Page 169

1   So do you see Ryan Mark's message at
2 13:13:35 PST on the first page?
3   A.   Yes, I see this message.
4   Q.   So Ryan says to Steven, wanted to
5 follow up on ████████
6   Do you see that?
7   A.   Yes, I see it.
8   Q.   Do you know what ████████s?
9   A.   I don't recall specifically.
10   Q.   Do you recall ████████ being datasets
11 that are publicly available?
12   A.   I have a vague recollection that it
13 was web crawl data, but I am not certain.
14   Q.   So by web crawl data, do you
15 understand that to be similar to or akin to
16 Common Crawl?
17   A.   Of a similar category to Common
18 Crawl.
19   Q.   So Ryan asks Steven if he wanted to
20 follow up on ████████ would that be fair to say?
21   A.   He asked if he would -- hi, Steven,
22 I wanted to follow up on ████████ Is that

Page 170

1 something that we are still interested in
2 pursuing.
3   That's what he says.
4   Q.   Steven Roller at 13:17:02, do you
5 see that message?
6   A.   Yes.
7   Q.   He says TBH, I think ████████
8 ████████
9   Do you see that?
10   A.   Yes.
11   Q.   TBH is -- do you understand that to
12 mean "to be honest"?
13   A.   Yes, I assume that's what he meant.
14   Q.   He says ████████ is a dead
15 end.
16   Do you understand that to mean that
17 Steven Roller believed that ████████ was not a
18 dataset to be used to pre-train the Llama?
19   A.   I am sorry. You are saying that
20 ████████ -- can you repeat that again?
21   Q.   Let me rephrase.
22   What do you understand him to mean

Page 171

1 when he says ████████ is a dead end?
2   MR. WEINSTEIN: Object to form.
3   A.   I am not sure I fully understand
4 what he means from this context.
5 BY MR. YOUNG:
6   Q.   You understand him to be discussing
7 whether or not ████████ would be appropriate to
8 use as pre-training material?
9   A.   It's likely that he is saying it's
10 not useful for some reason. I don't recall why
11 and it doesn't say here.
12   Q.   Later Mark says: Oh, because of the
13 quality, right?
14   A.   He does say that.
15   Q.   Do you understand him to mean the
16 quality of the dataset?
17   A.   I think he is asking if that's the
18 reason.
19   Q.   So Ryan is asking if Steven believes
20 that ████████ is, quote/unquote, a dead end, and
21 Ryan says because the quality, right?
22   A.   Yes, that's what I see here.

Page 172

1   Q.   Steven says: ████████
   ████████
   ████████
5   Do you see that?
6   A.   Yes, I see that.
7   Q.   And then he goes on: ████████
   ████████
9   Do you see that?
10   A.   Yes, I see that.
   ████████
   ████████
   ████████
   ████████
   ████
16   Q.   And then you respond to Steven,
17 right.
18   You say: ████████
   ████████
   ████████.
21   Right?
22   A.   Correct.

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 173

1   Q.   Do you recall what he referring to
2 when you said make a call soon?
3   A.   I don't specifically remember.
4   Q.   Were you referring to purchasing
5 data for use as pre-train material?
6   A.   From this context I think I was
7 referring to whether

9   Q.   You said Ryan is spending a ton of
10 time on this, right?
11   A.   Yes.
12   Q.   Do you remember what the this was
13 referring to?
14   A.   Unfortunately, no. I don't
15 remember.

18   Q.   And then Mr. Roller sent a
19 hyperlink. It appears to be an URL. Actually,
20 strike that.
21   Do you recognize what that URL is
22 that Steven Roller sends you?

Page 174

1   A.   I recognize -- I recognize the
2 domain name.

5   A.   Yes, that domain name looks
6 familiar.

9   A.   I think it is a provider of web
10 crawls.
11   Q.   Is it a provider of web crawls for
12 sale?
13   A.   I don't know their full business
14 model, but I believe you can purchase web crawls
15 from them.
16   Q.   Do you know if

20   Q.   Now, can you please turn to the next
21 page, last two, Bates Numbers 85.
22   A.   Yes.

Page 175

1   Q.   So looking at these message here,
2 were you still discussing

9   Q.   Now, do you see the reference to biz
10 dev team by you at 13:20:28?
11   A.   Yes.
12   Q.   Biz dev stands for business
13 development?
14   A.   That's correct.
15   Q.   What was biz dev's role when you
16 were referring to them in this message?
17   A.   My understanding of their role is
18 that they help us engage with potential vendors.

Page 176

7   Q.   Do you recall who was on the biz dev
8 team at this time?
9   A.   I believe -- I don't recall all the
10 names, but I believe I was working most closely
11 with Alex Boesenberg. I might be getting the
12 last name incorrect.
13   Q.   What about Amanda Kollet?
14   A.   I remember her. I don't remember
15 when she joined that team.
16   Q.   Was Mr. Boesenberg on biz dev team
17 as of December 2022?
18   A.   I believe so, but I -- I would have
19 to confirm that.
20   Q.   And what about Ms. Kollet, was she
21 on the biz dev team as of 2022?
22   A.   I don't recall.

44 (Pages 173 to 176)

Page 177

1   Q.   Okay. So Mr. Roller then says: I
2 mean before -- so Mr. Roller says, at 13:21:58,
3 he says: I mean before that, we should maybe
4 just sign up and investigate the data.
5   Do you see that?
6   A.   Yes, I see that.
7   Q.   Okay. And then you respond: Yeah.
8 Is there an easy way to do that.
9   Right?
10   A.   Yes, I see that.
11   Q.   What do you understand Mr. Roller to
12 mean by "investigate the data"?
13   A.   I think he's potentially referring
14 to a sample data that they gave.
15   Q.   Okay. And then Mr. Roller says:
16 Yeah. They literally have an API and I have a
17 credit card, LOL.
18   Right?
19   A.   Yes.
20   Q.   And then you respond: Nice. Yeah,
21 let's do that.
22   Right?

Page 178

1   A.   Yes.
2   Q.   Do you know if Mr. Roller ever
3 obtained the data he was referring to here?
4   A.   I do not know.
5   Q.   Okay. Do you know if he ever uses
6 credit card to purchase set data that he was
7 referring to here?
8   A.   I do not recall.
9   Q.   Okay. And then moving on. You
10 respond at 13:23:03, right? You say: Then we
11 have to make sure licenses work out, too.
12   Right?
13   A.   Yes.
14   Q.   What did you mean by "making sure
15 the licenses work out"?
16   A.   I meant that it's generally
17 acceptable for us to look at a dataset, but we
18 wouldn't want to train our models on a dataset
19 before going through a legal and privacy review.
20   Q.   To make sure that pre-training was
21 within the terms of the license.
22   Would that be fair to say?

Page 179

1   A.   I wasn't the one kind of parsing
2 the -- the licenses. I -- I don't know exactly
3 what to answer there.
4   Q.   But you were expressing concern
5 about making sure that you didn't run afoul of
6 the licenses or any licenses that you --
7   (Reporter clarification.)
8 BY MR. YOUNG:
9   Q.   -- run afoul of any licenses that
10 may come attached with the data that you were
11 potentially acquiring, right?
12   A.   I think -- I don't recall exactly
13 what I meant. But I think I was concerned with
14 making sure that we got our legal and privacy
15 team to review our data, as we always did.
16   Q.   Okay. So Mr. Roller goes on. He
17 says: There's a couple of vendors.

Page 180

22   Q.   Okay. Thank you.

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 181

9   A.   Yes, I see that.
10   **Q.   Okay. And then you say: I had to**
11   **be looped in again for more technical cues, but**
12   **that wasn't bad.**
13   **Right?**
14   A.   Yes.

21   **Q.   And this would be people like**
22   **Mr. Boesenberg or Ms. Kollet, right?**

Page 182

1   A.   That's correct.

14   **Q.   Now, if you can go back to the first**
15   **page of this document, ending Bates label**
16   **number 84.**
17   **In that first message, you share a**
18   **document, right?**
19   A.   It appears that there's an image.
20   **Q.   Yes, okay.**
21   MR. YOUNG: So I am going to pass
22   around a document bearing Bates label Meta

Page 183

1   Kadrey 00054287.
2   (Whereupon, Kambadur Exhibit 9,
3   Proposal Bates Number Meta Kadrey 00054287,
4   was marked for identification.)
5   MR. YOUNG: And this will be
6   Plaintiffs' -- marked as Plaintiffs' 9.
7   And marked as Plaintiffs' 9A will
8   also be the metadata associated with that
9   same document.
10   (Whereupon, Kambadur Exhibit 9A,
11   Metadata associated with Exhibit 9, was
12   marked for identification.)
13   BY MR. YOUNG:
14   **Q.   Let me know when you're ready.**
15   A.   Ready.
16   **Q.   Do you recognize this document?**
17   A.   Yes.
18   **Q.   What is this document?**
19   MR. WEINSTEIN: You're referring to
20   9 or 9A?
21   MR. YOUNG: Oh, sorry. Looking at
22   the doc -- Document 9, Plaintiffs' 9.

Page 184

1   MR. WEINSTEIN: Okay.
2   A.   It is a proposal for training
3   proportions for one of our language models.
4   BY MR. YOUNG:
5   **Q.   Do you know if this was the document**
6   **you shared in the chat in the document previously**
7   **marked as Plaintiffs' 8?**
8   A.   I don't recall.
9   **Q.   Okay. If you'll take a look at**
10   **Document 9A or Plaintiffs' 9A.**
11   **And then, do you see the field file**
12   **name, fifth from the bottom?**
13   A.   File path you mean?
14   **Q.   File name.**
15   A.   File name. Oh. Yeah, sorry. File
16   name, yes.
17   **Q.   Do you see that this file was**
18   **produced to us bearing the file -- with the file**
19   **name ending in .PNG?**
20   A.   Yes.
21   **Q.   Now, if you go back to**
22   **Plaintiffs' A -- Plaintiffs' Exhibit 8.**

46 (Pages 181 to 184)

Page 185

1   **Is that file name, that PNG file**
2   **path file name the same as the document that you**
3   **shared in your message at 7:22:16?**
4   A.   It's a really long file name.
5   **Q.   Yes.**
6   A.   Give me a second.
7   **Q.   No problem. Take your time.**
8   A.   Yes, it looks the same.
9   **Q.   Does that suggest to you that**
10   **these -- that this was the file that you share in**
11   **that message there?**
12   A.   It appears so, yes.
13   **Q.   So turning that back to**
14   **Plaintiffs' 9, if you look at the leftmost**
15   **column, can you describe for me what that column**
16   **is?**
17   A.   It's internal names of datasets.
18   **Q.   And what -- so at the very top it**
19   **says: To use in the large right, right?**
20   A.   Yes.
21   **Q.   Well, is -- what does that mean?**
22   A.   I'm not 100 percent sure. But based

Page 186

1   on the context and the date, I'm assuming this
2   may be -- yeah, it -- it could be one of our
3   large language model runs.
4   **Q.   Okay. If you look at the message**
5   **thread in which you attach this, this was a large**
6   **language model that you were running in**
7   **December 2022; is that correct?**
8   A.   That someone was presumably working
9   on in December 2022.
10   **Q.   Okay. Well, you shared this file,**
11   **right?**
12   A.   Yes.
13   **Q.   Okay. So how would you have gotten**
14   **this chart or this image?**
15   A.   How would I have gotten it? I --
16   perhaps someone shared the original file with me.
17   **Q.   Now, in that leftmost column, the**
18   **white column, those are the list of the datasets**
19   **that were used in this training model.**
20   **Would that be fair to say?**
21   A.   That -- that may be the case.
22   **Q.   All right. And let's go through**

Page 187

1   some of these datasets.
2   **So what is Stack Exchange?**
3   **(Reporter clarification.)**
4   A.   So I'm -- I want to clarify that I
5   didn't process any of this data. So I -- it
6   would all be a little bit assumptions or guesses
7   based on the names.
8   BY MR. YOUNG:
9   **Q.   Okay. Well, let me back up.**
10   **Would it be fair to say that this**
11   **likely came from someone you managed or**
12   **supervised?**
13   A.   Based on the context, I think it
14   actually came from the team -- that the -- the
15   people who were working on the data mix for
16   Llama 1, which were not people I directly
17   managed.
18   **Q.   So would it be fair to say that this**
19   **represents the, quote/unquote, data mix for what**
20   **would become Llama 1?**
21   A.   That, I am not certain. It's -- I
22   think it is a proposal of potential data mixes

Page 188

1   for that model.
2   **Q.   That would be what is reflected in**
3   **the column titled "Proposition 2," correct?**
4   A.   That's -- a proposition is what I am
5   guessing, proposal.
6   **Q.   So let's walk through some of these**
7   **datasets, what you understand them to be.** ▮
8   ▮ ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮
9   **Books3 is the dataset that came from EleutherAI**
10   **in the Pile, right?**
11   A.   I actually don't know the origin of
12   this dataset.
13   **Q.   But Books3 was the same dataset that**
14   **we were discussing earlier today, right?**
15   A.   Books3 has the same name. I don't
16   know if it was downloaded from Eleuther or the
17   Pile.
18   ▮ ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮
19   ▮ ▮▮▮▮▮▮▮▮▮▮▮
20   ▮ ▮▮▮▮▮▮▮▮▮▮▮▮▮▮
21   ▮ ▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮▮
22   ▮ ▮▮▮▮▮▮▮▮▮ ?

Page 189

1   A.   Yes, I see that.

[redacted]

8   Q.   Then arXiv? That's arXiv, right?
9   A.   That's -- we pronounce it archive,
10   if that's what you are asking.
11   Q.   Yes.
12   A.   Yes.
13   Q.   Because X is the Greek letter Chi?
14   A.   I'm not actually sure why, if that
15   sense.
16   Q.   Thank you.
17   And GitHub, GitHub is code, right?
18   A.   GitHub in this context likely refers
19   to the GitHub code repositories.
20   Q.   Have you ever used GitHub yourself?
21   A.   Yes.
22   Q.   Have you published code under

Page 190

1   GitHub?
2   A.   What do you mean by "published"?
3   Q.   Have you uploaded code onto a
4   repository on GitHub?
5   A.   Yes.
6   Q.   Did you upload codes subject to an
7   open source license on GitHub?
8   A.   Likely, yes.
9   Q.   Do you know if anyone has ever used
10   your code that you published on GitHub in
11   violation of the term of your license?
12   A.   I don't have particular knowledge of
13   it.
14   Q.   Would you be happy if someone did
15   that?
16   A.   Would I be happy if someone did
17   what?
18   Q.   Used your code in violation of its
19   license?
20   A.   It might depend on the
21   circumstances.
22   Q.   Like, for example, a lot of open

Page 191

1   source licenses and code that's been posed on
2   GitHub, for example, may have attribution terms.
3   Do you understand that?
4   A.   I understand what an attribution
5   term is, yes.
6   Q.   If someone were to take your code
7   that you published and did not provide
8   attribution, do you -- would you be happy with
9   that?
10   MR. WEINSTEIN: Object to form.
11   A.   I guess like, what do you mean would
12   I be happy with that? I don't precisely
13   understand.
14   BY MR. YOUNG:
15   Q.   Would you be pleased with that?
16   A.   I think it would likely depend on
17   the context of the use.
18   Q.   What is C4?
19   A.   C4 is, I believe, a version of
20   Common Crawl.
21   Q.   This is the colossal -- I forget
22   what it is myself. We will skip that.

Page 192

1   What is CCNet?
2   A.   I believe that's another version of
3   Common Crawl.
4   Q.   And Wikipedia, do you know what that
5   is?
6   A.   I assume in this context it refers
7   to documents scraped from the website Wikipedia.
8   Q.   Which comprised the trading dataset
9   used for pre-training this model or this proposed
10   model?
11   A.   This proposition is showing
12   Wikipedia as an option to include.
13   Q.   So the column labelled "Proposition
14   2," do you see that column?
15   A.   Yes.
16   Q.   It has a number of numbers in it.
17   A.   Yes.
18   Q.   Do you know what those numbers mean?
19   A.   Not completely.
20   Q.   Can you tell us what your
21   understanding is of those numbers?
22   A.   It seems to be some kind of a

48 (Pages 189 to 192)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 193

1 weighting.
2   Q.   And the column next to it is
3 "Percent," right?
4   A.   That's what it's labeled.
5   Q.   Does that column, "Percent," seem to
6 be also be indicative of some sort of weighting?
7   A.   Yes.
8   Q.   By "weighting," you mean weighting
9 of the dataset, right?
10   A.   I mean in the context of the first
11 column, I think it's simply a math tool to get
12 the other columns to add up right. In the
13 context of the percentage, I think it's what
14 percentage of the proposed training corpus row
15 would comprise.
16   Q.   Okay. In the "Number of Tokens"
17 column, that represents number of tokens came
18 from that dataset, right?
19   A.   I am not certain if this is raw
20 tokens or weighted tokens.
21   Q.   What is the difference?
22   A.   I am not sure if this is the

Page 194

1 original number of tokens or multiplied by the
2 number of epochs.
3   Q.   Can you explain for us again what
4 epochs mean?
5   A.   Epochs would be how many times we
6 provide the model this data as input.
7   Q.   So if you go back to the document
8 previously labeled as Plaintiffs' 8, message
9 thread in which you shared this document, so do
10 you see how you respond, say: Great. Thanks for
11 starting this, took a look and added a few
12 comments.
13   A.   I see that.
14   Q.   Do you recall what your comments
15 were?
16   A.   No. It's not really clear from the
17 context if the comments were on this document or
18 from some other document.
19   Q.   You then follow up: One last thing
20 we might want to/need to deal with, is discussing
21 co release with Genesis.
22   Right?

Page 195

1   A.   Yes, I said that.
2   Q.   Okay. With Genesis meaning what
3 would become the Llama 1 model?
4   A.   Yes.
5   Q.   Then you see a message 7:23:05: Got
6 the rest of this image. As we guessed, they
7 aren't using fiction or no OSS GitHub.
8   A.   Yes, I see that.
9   Q.   When they are saying "fiction," do
10 you understand that to mean Libgen fiction?
11   A.   No. I am not sure.
12   Q.   What do you -- thank you. Okay.
13 You can set that document aside.
14   Okay. Do you know who Sean Bell is?
15   A.   Yes.
16   Q.   Who is Sean Bell?
17   A.   He is currently the manager
18 supporting our data foundations team in gen AI.
19   Q.   Do you recall receiving a document
20 describing the mission and vision of the Data and
21 Knowledge Foundation?
22   A.   Yes.

Page 196

1   MR. YOUNG: I am going to mark
2 Plaintiffs' 10 document bearing Bates label
3 Meta Kadrey 00055255.
4   (Whereupon, Kambadur Exhibit 10,
5 Mission and vision statement Bates Number
6 Meta Kadrey 00055255, was marked for
7 identification.)
8 BY MR. YOUNG:
9   Q.   Ms. Kambadur, do you recognize this
10 document?
11   A.   I recognize that I saw at least one
12 version of this document previously.
13   Q.   Can you describe what this document
14 is?
15   A.   It's a mission and vision statement
16 for the data foundations team.
17   Q.   What is the data foundations team?
18   A.   It's a group that was formed early
19 2024 to focus on foundational model training.
20 It's a foundation model training data.
21   Q.   So by that do you mean pre-training
22 datasets or --

49 (Pages 193 to 196)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 197

1   A.   Pre-training datasets would be at
2   least some of what they focused on.
3   Q.   So are you a member of the data
4   foundations team?
5   A.   No.
6   Q.   Sean Bell heads that team?
7   A.   Sean Bell is managing that group,
8   yes.
9   Q.   Do you work closely with the data
10   foundations team?
11   A.   What do you mean by "work closely
12   with"?
13   Q.   How would you work with them at all?
14   A.   Do we work with them at all?
15   Q.   Yes.
16   A.   Yes, we do work with them.
17   Q.   How do you work with them?
18   A.   How do we work with them? In my
19   current role I am collaborating with them on
20   post-training data infrastructure.
21   Q.   Do you work with them at all for
22   pre-training data purposes?

Page 198

1   A.   Presently I am not working on
2   pre-training data but we occasionally get
3   information from them about pre-training data
4   that might inform our post-training efforts.
5   Q.   So do you see the section "Our
6   Vision"?
7   A.   Yes.
11   Do you see that?
12   A.   Yes.
13   Q.   It says:
15   Do you see that?
16   A.   I see that.
17   Q.   Do you understand that to be part of
18   the mission of the data foundations to this day?
19   A.   I'm not sure if they've updated
20   their mission since this document has been
21   written.
22   BY MR. YOUNG:

Page 199

Page 200

1   priority.

Page 201

1   A.   I see that statement.
2   **Q.   What is AGI?**
3   A.   I would assume it's referring --
4   referring to artificial general intelligence.
5   **Q.   Could you explain what artificial**
6   **general intelligence?**
7   A.   I'm not sure we have a precise
8   canonical definition of it. But my -- well, at
9   least one definition is, that AI has very broad
10   intelligence. Another definition might be that
11   it has intelligence beyond human intelligence or
12   matching human intelligence.
13   **Q.   Okay. The next bullet point says:**
     ██████████████████████████████████
15   **Do you see that?**
16   A.   Yes, I see that.
     █████████████████████████████
     ███████████
     ████████████████████████
     ████████████████████████████████
     ███████████████████████████████
     ███████████████████████

Page 202

1   **Q.   What do you understand -- what do**
2   **you understand saturated models to mean?**
3   A.   I -- I think it could be getting --
4   like, continuing to improve the performance of
5   the model and continuing to see, yeah, continued
6   metric improvements on our academic benchmarks,
7   for example.
     ████████████████████████████████
     ███████████████████████
     ██████████████████████████████
     ██████████████████████
     █████████████████████████████████
     ██████████████████████████
     ██████████████████████████████
     ███████████████████████████
     ████████████████████
18   **Q.   So it would be your -- it's -- your**
19   **reading of this document, it's a little bit**
20   **perhaps puffery?**
21   A.   Puffery. Sorry.
22   **Q.   Let me strike that. Let's move on**

Page 203

1   to the final bullet on this page.
     ███████████████████████████████
     ██████████████████████████████████
     █████████████████████████████████████
     ████████████████████████████
     ████████████████████████████████
     █████████████████████████
     █████████████████████████████████████
     ██████████████████████████
     ████████████████████████████████
     ██████████████████
13   **Q.   Okay. So is it part of your job --**
14   **you can set that document aside.**
15   **Is it your part of your side -- your**
16   **job as an AI researcher to research -- or read**
17   **and digest academic articles about developments**
18   **in generative AI?**
19   A.   I don't know if -- I don't know that
20   it was defined as a specific responsibility, but
21   it is -- I regularly read papers.
22   **Q.   And would it be fair to say that the**

Page 204

1   **papers that you read as part of your role as an**
2   **AI researcher or perhaps as someone interested in**
3   **AI, fairly sophisticated?**
4   **Would that be fair to say?**
5   A.   What do you mean that they are
6   fairly sophisticated?
7   **Q.   That's a fair question.**
8   **Would a layperson be able to**
9   **understand the articles that you read as part of**
10   **your role as an AI researcher?**
11   A.   That probably depends on the
12   layperson and the article.
13   **Q.   Fair enough.**
14   **So if a large language model were to**
15   **output a verbatim section of a work, could it be**
16   **inferred that that work was included in the**
17   **pre-training set?**
18   MR. WEINSTEIN: Object to form.
19   A.   Sorry. Can you repeat that again?
20   BY MR. YOUNG:
21   **Q.   Yeah.**
22   **So if an LLMs outputs a verbatim**

51 (Pages 201 to 204)
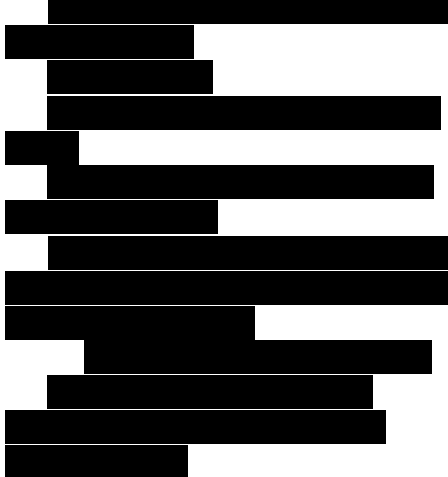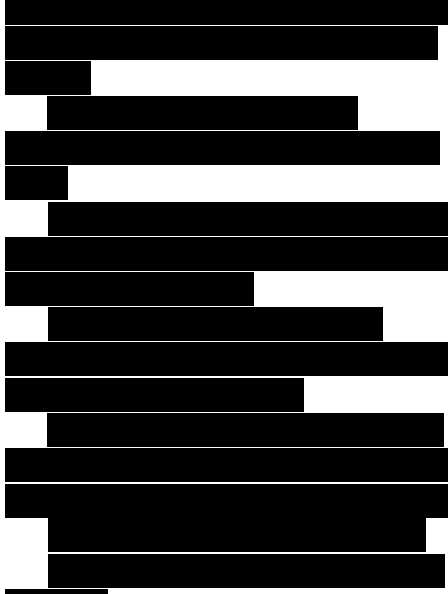
Page 205

1 section of a work, could you infer that that work
2 was included in the pre-training set?
3   A.   What do you mean by "work"?
4   Q.   So, for example, a piece of text or
5 a book.
6   A.   I don't think you would have
7 100 percent certainty that it -- it's in the
8 pre-training corpus and you certainly wouldn't
9 have any level of certainty of where that text
10 snippet came from originally.
11   Q.   But certainly a possibility, right?
12   MR. WEINSTEIN: Object to form.
13   A.   Sorry. Is what a possibility?
14 BY MR. YOUNG:
15   Q.   That the work that was being
16 regurgitated verbatim was included in the
17 pre-training set.
18   A.   If you find a text output of a model
19 that it's possible that it could match something
20 provided in an input. But we generally see that
21 in, like -- in very limited scenarios, very small
22 percentages and very small spans of text output.

Page 206

1   MR. YOUNG: All right. Let's mark
2 as Plaintiffs' 11, a document bearing Bates
3 Number Meta Kadrey 00054433.
4   (Whereupon, Kambadur Exhibit 11,
5 Chat message Bates Number Meta Kadrey
6 00054433, was marked for identification.)
7   MR. YOUNG: So I think we've been
8 going little less than an hour. So let's
9 finish this document and we can take a break.
10 BY MR. YOUNG:
11   Q.   Do you recognize this document,
12 Ms. Kambadur?
13   A.   I don't particularly recognize the
14 conversation, but it -- it looks like a chat
15 between me and a teammate.
16   Q.   This is a Workplace -- does this
17 appear to be a Workplace Chat between you and
18 Mr. Esiobu that took place on or around April 5,
19 2023, in the regular course of business?
20   A.   Yes, it appears to be.
21   Q.   Okay. And you and Mr. -- then
22 this -- and April 5, 2023, was that about the

Page 207

1 time the Llama 1 model or shortly after the
2 Llama 1 model was released?
3   A.   I don't recall the exact date of
4 Llama 1 release. That was -- was that February?
5   Q.   Yes.
6   A.   So maybe this is a couple months
7 later.
8   Q.   So at this point, Llama was already
9 available, right?
10   A.   Yes.
11   Q.   And David's first message goes:
12
13   Do you see that?
14   A.   Yes, I see that.
15   Q.   Okay. And then he shares a file
16 with you, right?
17   A.   Looks like there's an image.
18   Q.   Yeah. It's a PNG file?
19   A.   Yes.
20   MR. YOUNG: So I am going to mark as
21 Plaintiffs' 12, a document bearing Bates
22 label Meta Kadrey 00054435, and also as 12A,

Page 208

1 the metadata associated with that document.
2   (Whereupon, Kambadur Exhibit 12,
3 File Bates Number Meta Kadrey 00054435, was
4 marked for identification.)
5   (Whereupon, Kambadur Exhibit 12A,
6 Metadata for Exhibit 12, was marked for
7 identification.)
8 BY MR. YOUNG:
9   Q.   Do you recognize the document that's
10 been marked as Plaintiff's 12?
11   A.   I have a vague recollection.
12   Q.   Do you recall this as the file -- do
13 you recall this is the file that Mr. Esiobu
14 shared with you on this message thread? So
15 what's place --
16   A.   I don't specifically recall, but.
17   Q.   If you look at what's been marked as
18 Plaintiff's 12A, do you see the file name row?
19   A.   Yes.
20   Q.   Is that the same file name as the
21 file that Mr. Esiobu shared with you in that
22 Workplace -- Workplace thread?

52 (Pages 205 to 208)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 209

1   A.   It appears to be, yes.
2   **Q.   Does that make you believe that this**
3   **is the same file that Mr. Esiobu shared with you**
4   **on April -- in April of 2023 when you had this**
5   **conversation with him?**
6   A.   It appears to be, yes.
7   **Q.   So Plaintiffs' 12, can you describe**
8   **for us what this document is?**
9   A.   It looks like some example inputs
10 and outputs of one of our models.
11   **Q.   The Chat Llama, that is one of the**
12 **models that you had, right, that -- excuse me.**
13 **Let me rephrase that.**
14   **Chat Llama is one of the Llama**
15 **models that Meta produced, right?**
16   A.   What do you mean by "produced"?
17   **Q.   Excuse me, let me try that again.**

Page 210

1   **Q.   Okay. So if you look at these**
2   **inputs and outputs, right, do you see that there**
3   **are some figures associated with each input and**
4   **output, right, little pictograms?**
5   A.   Yes. Do you mean the little image?
6   **Q.   One of them is kind of a rough**
7   **approximation of a person.**
8   **Do you see that?**
9   A.   Yes, I see that.
10   **Q.   And the other one is kind of a rough**
11 **approximation of what appears to be a robot face?**
12   A.   Yes.
13   **Q.   So the one that looks like a person,**
14 **is that -- would that -- would it be fair to say**
15 **that those are the inputs?**
16   A.   Yes. Those look like the user input
17 prompts.
18   **Q.   The ones with the robots, those are**
19 **the output -- excuse me.**
20   **The ones with the robot face, those**
21 **are the outputs that chat Llama bot -- or the**
22 **chat Llama model is outputting, correct?**

Page 211

Page 212

www.DigitalEvidenceGroup.com Digital Evidence Group C'rt 2024   202-232-0646

Page 213

6   (Court reporter clarification.)
7   BY MR. YOUNG:
8   Q.   Goodreads.com, G-O-O-D-R-E-A-D-S dot
9   com. Do you see that message at 8:49:09 PDT?
10   A.   Yes, I see the message.
11   Q.   Do you know what goodreads.com is?
12   A.   It's a website for writing and
13   discussing books.
14   Q.   He says, "it has pulled quotes from
15   many books in that one match"?
16   A.   Yes, that's what he said.
17   Q.   He provides an -- appears to be an
18   URL to a web page, right?
19   A.   Yes. It looks like an URL to a web
20   page.
21   Q.   And you will have a couple messages.
22   Then David responds again. Oh, I guess it is

Page 214

1   actually, and then provides a link.
2   Do you see that?
3   A.   I see that link.
4   Q.   Okay. Do you recognize that link?
5   A.   I don't particularly recognize the
6   link.
7   Q.   Okay. Do you understand it to be a
8   link to an internal drive or internal file within
9   Meta's file system?
10   A.   I think it is.
15   Q.   What is an epub?
16   A.   I think that's maybe a publishing
17   format.
18   Q.   Do you understand that to be a file
19   format for electronic books?
20   A.   I think that's correct.
21   Q.   Dot TXT, that's means it's a text
22   file, correct?

Page 215

Page 216

17   BY MR. YOUNG:
18   Q.   Can you -- do you see at the bottom
19   David Esiobu, on the same document, same page,
20   has a series of three messages at the very
21   bottom, right? Right before that you ask -- you
22   have two message above that. Do you see that?

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 217

1   **The first bearing time stamp 9:13:45 PDT.**
2   A.   Yes, I see that.
3   **Q.   You ask: Is there a way for me to**
4   **see the file in this dataset BTW?**
5   **Right?**
6   A.   Yes.

Page 218

18   **Q.   You can tell I am not a code person.**
19   **I don't know back slashes.**

Page 219

1   MR. WEINSTEIN: Object to form.
2   A.   Sorry. Can you -- does that path
3   say --
4   BY MR. YOUNG:
5   **Q.   Let me try that again.**

16   BY MR. YOUNG:

Page 220

17   **Q.   Now, do you see your message at**
18   **9:03:23?**
19   A.   Yes, I see me -- yeah, sorry. Yes.
20   **Q.   What did you say at 9:20 -- 9:03:23?**
21   A.   You want me to read the message?
22   **Q.   Yes, please.**

55 (Pages 217 to 220)

Page 221

10   Q.   Is it because -- do you understand
11 "Harry Potter and the Sorcerer's Stone" to be
12 within copyright?
13   MR. WEINSTEIN: Object to form.
14   A.   To be within copyright? Can you
15 define that for me? Sorry, I'm not.
16 BY MR. YOUNG:
17   Q.   Let me -- let me -- let me rephrase.
18   Do you understand "Harry Potter and
19 the Sorcerer's Stone" to be copyrighted?
20   A.   I assume it has some kind of
21 copyright.
22   Q.   Well, were you expressing concern

Page 222

1 because the Llama model was regurgitating
2 portions of a copyrighted book?
3   MR. WEINSTEIN: Object to form.
4   A.   I was expressing some concern
5 because I personally did not know Harry Potter
6 was in the Books3 datasets. But I think from
7 this conversation and this screenshot, we don't
8 actually know that
10 BY MR. YOUNG:
17   Q.   Now, just a couple more questions
18 and I think we can take a break.

Page 223

10   Q.   Okay. By the way, do you see that
11 last Google Doc that David Esiobu shared at the
12 very bottom?
13   A.   Yes.
14   Q.   Do you know if that document is
15 still alive?
16   A.   I assume it is.
17   Q.   Do you know if we can access it
18 today?
19   MR. WEINSTEIN: Object to form.
20   A.   I don't know how it -- I don't know
21 what was provided to you already by our legal
22 team.

Page 224

1   MR. YOUNG: Counsel, if that
2 document wasn't provided, I think we're going
3 to make a request that be provided.
4   MS. STAMESHKIN: Noted.
5   MR. YOUNG: Thank you.
6   (Whereupon, a request for Production
7 of Google document shared by David Esiobu was
8 made.)
9 BY MR. YOUNG:

Page 225

19   MR. WEINSTEIN: Object to form.
20   A.   I -- I mean, I don't know what is
21 the first line. It could be.
22   MR. YOUNG: I'm done with this

Page 226

1   document. So this I think we're -- this
2   might be a good time for a break.
3   MR. WEINSTEIN: Sure.
4   THE VIDEOGRAPHER: The time right
5   now is 2:47 p.m. We are off the record.
6   (Whereupon, a brief recess was
7   taken.)
8   THE VIDEOGRAPHER: The time right
9   now is 3:02 p.m. We are back on the record.
10 BY MR. YOUNG:
11   Q.   Do you know who Jacob Zhou is?
12   A.   Yes.
13   Q.   Who is Jacob Zhou?
14   A.   He's a data engineer.
15   Q.   And what about Susan Zhang?
16   A.   Susan was previously a research
17 engineer at Meta.
18   Q.   So she -- she's no longer employed
19 at Meta?
20   A.   No.
21   Q.   Do you know where she works now?
22   A.   I believe she's at Google.

Page 227

1   Q.   And when did she leave Meta for
2   Google?
3   A.   I don't believe she went directly to
4   Google.
5   Q.   Do you know where she went after she
6   left Meta?
7   A.   No. I remember her talking about
8   potentially working on a start-up, but I don't
9   know exactly where she landed.
10   MR. YOUNG: Okay. I'm going to mark
11 Plaintiffs' 13, a document bearing Bates
12 label Meta Kadrey 00054894.
13   (Whereupon, Kambadur Exhibit 13,
14 Workplace Chat Bates Number Meta Kadrey
15 00054894, was marked for identification.)
16 BY MR. YOUNG:
17   Q.   Do you recognize this document,
18 Ms. Kambadur?
19   A.   I don't. I haven't reviewed it yet.
20 Give me a second.
21   Q.   Okay. Take your time. Please let
22 me know when you're done reviewing the document.

Page 228

1   A.   Yes, I'm done reviewing the
2   document.
3   Q.   Excellent. Do you see your name in
4   the "To" field of this document?
5   A.   Yes.
6   Q.   Do you remember receiving this
7   document?
8   A.   I don't recall.
9   Q.   Okay. Does this document appear to
10 be a Workplace Chat on or around October 11,
11 2022, in which you were a participant in the
12 thread?
13   A.   Yes, it does.
14   Q.   And this was a Workplace Chat that
15 was generated in the regular course of business,
16 correct?
17   A.   Yes, it looks like it.
18   Q.   And the first message is from
19 Jacob Zhou, who we were just talking about,
20 correct?
21   A.   Correct.
22   Q.   And he is still presently employed

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 229

1   at Meta, correct?
2   A.   I believe so.

███████████████████
███████████████████████
█████████████████████████████
██████████████████████
████████
███████████████████████
██████████████████████████
████████████████████
███████████████████
███████████████████████████
██████████████████

14   Q.   Okay. So Jacob shares two files
15   with you, correct?
16   Do you see that?
17   A.   I see two PNG.
18   Q.   At 5:35:36 PDT, correct?
19   A.   Correct.
20   MR. YOUNG: So I am going to mark as
21   Plaintiffs' 14, a document bearing Bates
22   label Meta Kadrey 000054896.

Page 230

1   (Whereupon, Kambadur Exhibit 14,
2   List of file name strings and number pairs
3   Bates Number Meta Kadrey 000056896, was
4   marked for identification.)
5   MR. YOUNG: And as 14A, metadata
6   associated with that document.
7   I am also marking Plaintiffs' 15 --
8   (Whereupon, Kambadur Exhibit 14A,
9   Metadata for Exhibit 14, was marked for
10   identification.)
11   MR. YOUNG: Also marked
12   Plaintiffs' 15, a document bearing Bates
13   label Meta Kadrey 000054897. And 15A is the
14   metadata associated with that document.
15   (Whereupon, Kambadur Exhibit 15, PNG
16   file Bates Number Meta Kadrey 000054897, was
17   marked for identification.)
18   (Whereupon, Kambadur Exhibit 15A,
19   Metadata for Exhibit 15, was marked for
20   identification.)
21   BY MR. YOUNG:
22   Q.   Let me know when you've reviewed the

Page 231

1   documents.
2   Okay. So why don't you take a look
3   at the document marked -- marked Plaintiffs' 14.
4   Do you recognize this document?
5   A.   No.
6   Q.   Take a look at the document
7   marked 14A.
8   Do you see the file name?
9   A.   Yes. I see the file name.
10   Q.   It's a .PNG file?
11   A.   Yes.
12   Q.   Does that PNG file match the top
13   document that Jacob shares at 5:35:36 PDT?
14   A.   Yes, it appears to match.
15   Q.   Does that tell you that this is most
16   likely the document that Jacob shares in that
17   first link there in the message bearing time
18   stamp 5:35:36 PDT?
19   A.   That would be my assumption.
20   Q.   Can you describe this document for
21   us, the document marked as Plaintiffs' 14?
22   A.   It's a list of what appears to be

Page 232

1   file name strings and number pairs.
2   Q.   And some of the files -- some of the
3   strings have red highlights on them, right?
4   A.   Yes.
5   Q.   Okay. And each of the strings has
6   an element that is a .epoch.txt, correct?
7   A.   I have not looked at all of them,
8   but it looks -- yeah, it looks like it was most
9   of them, maybe all of them.

████████████████████████
███████████████████
██████████████████████████
████████████████████████████
█████████████████████
███████████████████
██████████████████████████
████████████████████
██████████████████████████████
███████████████████
█████████████████████████
██████████████████████

Page 233

3   Q.   Can you tell us what the decimal
4 strings mean?
5   A. I have no idea.
6   Q. So if we go back to what was marked
7 as Plaintiffs' 13, the message string, that first
8 message that Jacob sends at 5:35:36, the one with
9 the text, do you recall that he says that he

11   A. I see that in the document.

15   Do you see that?
16   A. Yes, I see that.
17   Q. Do you understand what he means by

19   A. I am not certain.
20   Q. Do you have a high level of
21 understanding what he might mean?
22   A. You want me to guess what he meant?

Page 234

1   Q. Based on the context of this
2 message.
3   MR. WEINSTEIN: Object to form.
4   You can answer.
5   A. I don't know if there's enough
6 context in this message to answer for sure, but
7 it could stand for threshold, the THRS.
8 BY MR. YOUNG:
9   Q. What would that mean, "threshold"?
10   MR. WEINSTEIN: Object to form.
11   A. You mean what does the word
12 "threshold" mean? What do you mean?
13 BY MR. YOUNG:
14   Q. So he's saying   right?
15   MR. WEINSTEIN: Object to form.
16   A. Yes. He says   which may
17 mean threshold.
18 BY MR. YOUNG:
19   Q. What do you understand
20 to mean?
21   A. That --

Page 235

2   Q. Value of what?
3   A. I am not certain.
4   MR. WEINSTEIN: Object to form.
5 BY MR. YOUNG:
6   Q. So if you take a look at -- go back
7 to Plaintiffs' 14, the document with the list of
8 strings with the red lines, do you see how every
9 item with the red line underneath is
10   A. Yes. That appears to be the case.
11   Q. So based on that message in
12 Plaintiffs' 13 that Jacob sent, do you understand
13 that   ?
14   MR. WEINSTEIN: Object to form.
15   A. He is saying   . I
16 don't know from what or by what standard.
17 BY MR. YOUNG:
18   Q. Okay. Do you see his message at
19 11:10:57 PDT? Let me know when you're there.
20   A. Yes, I see this message.
21   Q. He says: Hi, team. About content

Page 236

1 work a few months ago.

21   Q. Okay. Do you know what he might
22 mean by content distribution?

Page 237

1 A. I can -- I don't recall particularly
2 but I could hazard a guess based on the context,
3 I suppose.
4   Q. That's all right. Let's move on.
5   He says: TLDR.
6   Do you understand to mean too long,
7 didn't read?
8 A. Yes.
9   Q. It goes too long, didn't read and
10 then he's basically summarizing a message, right?
11   Do you understand TLDR to
12 essentially be summarizations?
13 A. I wouldn't say it's not necessarily
14 summarizing a message. It might just be giving a
15 briefer description. That's not actually a
16 summary of -- there may not be a longer version
17 that exists. He may be just be giving a brief
18 description.
19   Q. Okay. He says: I basically
20 reproduce the topic analysis on the Pile paper.
21 He use Pile.CC to train an LDA an inference on
22 other corpuses.

Page 238

1   What is an LDA?
2 A. I forget precisely what it stands
3 for. I have a guess but I don't want to say it
4 wrong.
5   Q. Can you generally describe what an
6 LDA is?
7   MR. WEINSTEIN: Object to form.
8 A. In this case he is talking about an
9 unsupervised clustering algorithm, I believe.
10   Q. Can you explain what an unsupervised
11 clustering is?
12 A. This means that you are trying to
13 group data into clusters without necessarily
14 having ground truth categories or labels defined
15 ahead of time.
16   Q. I am sorry. Did you say round
17 truth?
18 A. Ground truth.
19   Q. Got it. Thank you.
20   What is ground truth?
21 A. What I meant was -- generally when
22 we say ground truth, we mean a human-verified

Page 239

1 label, but it could be some other verification
2 method.
3   Q. So do you know what the Pile.CC --
4 do you know who he is referring to?
5 A. I can make a guess, but I don't know
6 for sure.
7   Q. So is that a dataset contained
8 within the Pile?
9   MR. WEINSTEIN: Object to form.
10 A. I don't know. I assume it has some
11 relation to Pile.
12 BY MR. YOUNG:
13   Q. So do you see Susan Zhang's responds
14 to him at 16:03:03 PDT, right?
15 A. Yes.
16   Q. So she says: Whoa, how did you have
17 Books3 in this analysis already back then?
18   Do you see that?
19 A. Yes.
20   Q. Okay. She is referring to an
21 analysis that Jacob links to there, right, at
22 GitHub?

Page 240

1 A. I assume so.
2   Q. Did you examine Jacob's analysis
3 yourself?
4 A. I don't recall.
5   Q. But Susan is expressing surprise
6 because it looks like Jacob used Books3 in that
7 analysis, correct?
8 A. I am not sure. I could guess based
9 on the use of "Whoa," but I don't want to say
10 what Susan's emotions were at the time.
11   Q. Based on the context of this
12 sentence, it appears that Jacob, at least, was in
13 possession of Books3 dataset before October 11,
14 2022, correct?
15   MR. WEINSTEIN: Object to form.
16 A. From this message, I think there is
17 some reason that Susan is guessing that there is
18 something related to Books3, but I didn't look --
19 I don't recall looking at the actual notebook
20 link, so I am not sure how she is putting those
21 pieces together.
22 BY MR. YOUNG:

60 (Pages 237 to 240)

Page 241

1   Q.   Do you see Jacob's message at
2   17:48:56 PDT?
3   A.   Yes.
4   Q.   And five numbers there, right?
5   A.   Yes.
6   Q.   What does Jacob say under 1?
7   A.   You want me to -- just the text.
8   Q.   Yes.
9   A.   Books3 is included in the Pile, so I
10   took it for free.
11   Q.   He has little tilde at the very end,
12   right, a little squiggly?
13   A.   Yes.
14   Q.   Do you understand that squiggly to
15   have any meaning?
16   A.   If it does, I don't know what
17   meaning.
18   Q.   Do you know if squigglies convey
19   playfulness or perhaps -- do you know if
20   squiggles convey playfulness perhaps?
21   A.   I've never seen that before, but
22   it's possible. I don't recall seeing it before

Page 242

1   at least.
2   Q.   Reading this message, do you
3   understand that Jacob to have taken Books3
4   without paying anything for it?
5   MR. WEINSTEIN: Object to form.
6   A.   It's hard to precisely interpret or
7   parse this sentence. For example, it doesn't
8   tell us if he's actually directly downloaded
9   anything.
10   BY MR. YOUNG:
11   Q.   What do you understand the phrase "I
12   took it for free" to mean?
13   MR. WEINSTEIN: Object to form.
14   A.   I'm not certain what it means. It's
15   not necessarily referring to cost. It could also
16   be referring to a lack of extra work.
17   BY MR. YOUNG:
18   Q.   Okay. You can set that aside --
19   document aside.
20   MR. YOUNG: Okay. We are going to
21   mark as Plaintiffs' 16, a document bearing
22   Bates label Meta Kadrey 00074155.

Page 243

1   (Whereupon, Kambadur Exhibit 16,
2   Workplace Chat Bates Number Meta Kadrey
3   00074155, was marked for identification.)
4   BY MR. YOUNG:
5   Q.   Why don't you take a moment to
6   review the document and let me know when you're
7   ready?
8   A.   Yes, I reviewed the document.
9   Q.   Does this document appear to be a
10   Workplace thread -- a Workplace Chat between
11   yourself and others on or around March 29, 2023?
12   A.   Yes, it does.
13   Q.   And this was created in the regular
14   course of business?
15   A.   It appears to be.
16   Q.   And let's go over some of the
17   participants in this thread we haven't spoken
18   about before.
19   Who is Sergey Edunov?
20   A.   He's a research director at Meta.
21   Q.   And was he a research director at
22   Meta as of March 29, 2023?

Page 244

1   A.   Yes.
2   Q.   And what about Moya Chen, who is
3   that?
4   A.   She is a research engineer.
5   Q.   Is she still employed at Meta?
6   A.   I believe she's returning to
7   employment at Meta after a break.
8   Q.   Okay. Thank you.
9   And Guillaume Lample, we've -- we've
10   spoken about him before, correct?
11   A.   Have we spoken about
12   Guillaume Lample?
13   Q.   Yeah.
14   A.   Yes.
15   Q.   Same individual?
16   A.   Yes.
17   Q.   Okay. So do you see Moya Chen's
18   doc -- message at 7:05:55 PDT?
19   She asks Guillaume: Do you have the
20   list of datasets and the path location written
21   out somewhere?
22   Do you see that?

Page 245

1   A. Yes, I see that.
2   Q. Okay. And then Guillaume provides a
3   link, right?
4   A. Yes.
5   Q. Now, do you know what list of
6   datasets Moya is referring to?
7   A. I don't recall.
8   Q. Okay. So this document was -- was
9   created on or about March 29, 2023.
10   Do you think this might have
11   something to do with either Llama 1 or Llama 2?
12   A. I don't know we can assume the
13   document was created then.
14   Q. Okay.
15   A. It looks like it was shared then.
16   Q. Okay. And then Guillaume, right,
17   Guillaume seems -- shares a document and then
18   points Moya to where she can find the information
19   she requests, right?
20   A. He says: It's in new run V2, which
21   could be referring to some part of that sheet.
22   (Reporter clarification.)

Page 246

1   THE WITNESS: V2, as in Victor, yes,
2   version, most likely.
3   BY MR. YOUNG:
4   Q. And -- sorry.
5   And then you respond, correct?
6   A. Yes.
7   Q. You say: Thanks for sharing. Can
8   we get the exact locations/versions of the data
9   in the sheet, too.
10   Right?
11   A. Yes.
12   Q. Okay. So you looked at that
13   document that Guillaume shared, right?
14   A. Based on this context, presumably,
15   yes.
16   Q. Okay. Now, do you remember what
17   list of datasets Moya was looking at?
18   A. No.
19   Q. Okay. Now, you make another request
20   of Guillaume, right? You said: We need for date
21   eval/keeping up --
22   Well, strike that.

Page 247

1   Now, Guillaume then shares a PNG
2   file, right? Do you see that at 7:16:34?
3   A. Yes, I see that.
4   MR. YOUNG: Now, I'm going to mark
5   as Plaintiffs' 17.
6   (Whereupon, Kambadur Exhibit 17, PNG
7   file shared in Exhibit 16 Bates Number Meta
8   Kadrey 3000741456, was marked for
9   identification.)
10   MR. YOUNG: Well, I'm sorry. Yep,
11   Plaintiffs' 17, a document bearing Bates
12   label Meta Kadrey 00074156, as well as 17A,
13   which is metadata associated with that
14   document.
15   BY MR. YOUNG:
16   Q. I want you to take a moment to
17   review that document. Let me know when you're
18   ready.
19   (Whereupon, Kambadur Exhibit 17A,
20   Metadata associated with Exhibit 17, was
21   marked for identification.)
22   A. Yes, I've looked at the document.

Page 248

1   BY MR. YOUNG:
2   Q. Now, do you recognize this document?
3   A. I've seen, I think, parts of the
4   sheet at least or versions of the sheet.
5   Q. Okay. I'm sorry. We're referring
6   to the document marked as Plaintiffs' 17, right?
7   That's the one you were looking at?
8   A. Yes.
9   Q. Okay. Is this the -- is this the
10   document that Guillaume shares with you in the
11   message thread marked as Plaintiffs' 16?
12   A. I don't remember. It's been a year
13   and a half.
14   Q. All right. Can you look at what's
15   been marked Plaintiffs' 17A, which is the
16   metadata for Plaintiffs' 17. And can you please
17   look at the field labeled as "File name"?
18   Do you see that?
19   A. Yes.
20   Q. And that file name is a PNG file?
21   A. Correct.
22   Q. And is that PNG file the same as the

Page 249

1 one that Guillaume shares at 7:16:34 PTD on
2 March 29, 2023?
3   A.   It appears to be.
4   Q.   Okay. Does that suggest to you that
5 that file, what is Plaintiffs' 17, is the same
6 file that Guillaume shares with you on March 29,
7 2023?
8   A.   That would be my assumption.
9   Q.   So returning to Plaintiffs' 17, do
10 you understand this to be a screenshot of a
11 larger sheet or Excel sheet?
12   A.   That's what it looks like.
13   Q.   Okay. And what do you understand
14 this section of the sheet to represent?
15   A.   It looks like a proposal for a data
16 mix for a version of a model.
17   Q.   Based on the timing of the message
18 thread, do you think this might be an early
19 version of Llama 2?
20   A.   What do you mean by "early version"?
21   Q.   Do you think this document -- let me
22 rephrase that.

Page 251

1 could refer to datasets, yes.
2   Q.   Okay. And we've seen some of these
3 names before, right?
4   A.   Yes.
5   Q.   Would the fact that these datasets
6 were listed on multiple propositions mean that
7 the -- these are the same datasets being used?
8   MR. WEINSTEIN: Object to form.
9   A.   Sorry. Which -- the same across
10 what documents or what references?
11 BY MR. YOUNG:
12   Q.   So we've seen, for example, Stack
13 Exchange listed on multiple documents today,
14 right?
15   A.   Yes.
16   Q.   Does that mean -- for multiple
17 experiments, correct?
18   A.   We've seen it on multiple documents.

Page 250

15 this actually -- actually become a trained model.
16 BY MR. YOUNG:
17   Q.   Okay. So under the column named --
18 labelled "New Run V2," do you see that?
19   A.   Yes.
20   Q.   Underneath, are those a list of
21 datasets?
22   A.   They look like a list of names that

Page 252

17   Q.   Okay. And arXiv, we've seen that
18 one before, correct?
19   A.   Yes, we've seen that name.
20   Q.   Okay. And I don't think we asked,
21 but arXiv, do you -- what do you understand arXiv
22 to be?

Page 253

1   A.   In this case, a dataset containing
2   academic articles.
3   Q.   **Including potentially the Llama 1**
4   **and Llama 2 papers?**
5   A.   I couldn't be sure about that. I
6   don't know when this was pulled.
7   Q.   **But this represents a scraping on**
8   **the arXiv model at some point in time, correct?**
9   A.   Of the arXiv model?
10   Q.   **Excuse me. The arXiv website?**
11   MR. WEINSTEIN: Object to form.
12   A.   I don't know how the dataset was
13   constructed. I'm aware that we had a dataset
14   named arXiv which contained articles.
15   BY MR. YOUNG:
16   Q.   **Which you understood to have -- at**
17   **some point been hosted on the arXiv website?**
18   A.   That was and is my assumption.
19   Q.   **I am sorry. Was?**
20   A.   Was and is my assumption, yes.
21   Q.   **Got it. Thank you.**
22   **GitHub, did you understand GitHub to**

Page 254

1   **be a scraping of the code repositories on GitHub**
2   **at some point in time?**
3   A.   Again, I don't know how the dataset
4   was created.
5   Q.   **Did you have an assumption about**
6   **what was contained in that dataset?**
7   A.   I understand that it contains code
8   data.
9   Q.   **Publicly available code data scraped**
10   **from public repositories online at GitHub?**
11   MR. WEINSTEIN: Object to form.
12   A.   I don't know our definition of
13   scraping nor do I know how this data was
14   collected.
15   BY MR. YOUNG:
16   Q.   **Well, C4 EN, do you know what that**
17   **dataset is?**
18   A.   I have a guess.
19   Q.   **That was the Common Crawl subset or**
20   **the Common Crawl dataset we were referring to**
21   **earlier?**
22   A.   That is a version of Common Crawl,

Page 255

1   and, I am assuming that EN refers to English
2   language.
3   Q.   **Okay. And there are a number of**
4   **other C4 items, right?**
5   A.   Correct.
6   Q.   **Do you understand those to refer to**
7   **datasets in certain languages?**
8   A.   That's my assumption based on these
9   two characteristics being common language codes.
10   Q.   **So FR would be French?**
11   A.   That would make sense.
12   Q.   **Would ES be Spanish?**
13   A.   It matches the Spanish language
14   code, so that would make sense.
15   Q.   **Would DE be German?**
16   A.   It sounds right, but I'm not sure.
17   Q.   **What is PT?**
18   A.   I am not 100 percent sure, but I
19   could guess.
20   Q.   **Portuguese?**
21   A.   Yes, that would be my guess.
22   Q.   **What about RU?**

Page 256

1   A.   My guess would be Russian.
2   Q.   **IT would be Italian?**
3   A.   That's what I would think.
4   Q.   **What is ZH?**
5   A.   I am not sure.
6   Q.   **Would it be some -- it would be a**
7   **dataset in whatever language Z stood for?**
8   A.   Yes, that's my assumption.
9   Q.   **And Wikipedia, we've seen that one**
10   **as well, right?**
11   A.   We have seen that name.
12   Q.   **Do you understand that to be**
13   **comprised of articles that were hosted on**
14   **Wikipedia?**
15   A.   That would be my assumption but
16   again, I didn't directly examine our construct of
17   this dataset.
18   Q.   **So there are certain columns in**
19   **orange.**
20   **Do you see those?**
21   A.   Yes.
22   Q.   **Now, we've seen similar labels in**

64 (Pages 253 to 256)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 257

1 column before, correct?
2   A.   We have seen some similar labels,
3 yes.
4   Q.   Can you describe what the column
5 labelled "Proposition 1 Weights" is?
6   A.   I don't actually know the
7 mathematically relation to the other columns, but
8 I assume it's some kind of weighting system.
9   Q.   Can you describe for us what it
10 would be mean if there's a higher weight number
11 versus a lower weight number?
12   A.   I would have to look. Yes, I would
13 have to compute the relationship between these
14 columns. I am not actually sure.
15   Q.   Okay. Thank you.
16   Column of number of tokens, right,
17 that's -- B stands for billions; is that right?
18   A.   The B likely stands for billions,
19 but it doesn't totally make sense to me.
20   Q.   It's all right.
21   In the last column on the very left,
22 you see number of epochs, right?

Page 258

1   A.   Yes. That's what it says, number
2 assigned, epochs.
3   Q.   Why are there more epochs on certain
4 datasets than others?
5   A.   In this data mix proposal
6 particularly, I am not sure.
7   Q.   Is that something that someone who
8 experimenting can set the number of epochs?
9   A.   We can change the number of epochs
10 as a variable in the data mix, yes.
17   Q.   Would a possible reason be to
18 increase memorization of that dataset?
19   MR. WEINSTEIN: Object to form.
20   A.   What do you mean by "memorization"?
21 Can we define that?
22 BY MR. YOUNG:

Page 259

1   Q.   Do you recall earlier today we had a
2 discussion about memorization and desirable
3 memorization versus undesirable memorization? Do
4 you recall that?
5   A.   I recall that conversation.
6   Q.   Would increasing desirable
7 memorization be a reason you would increase the
8 epochs in certain datasets?
9   A.   I am still not sure I understand
10 exactly what desirable memorization is.
11   Q.   It's okay. So you said that one of
12 the reasons that you may increase the number of
13 epochs is to increase the relative proportion of
14 that dataset in the model.
15   Do you recall that?
16   A.   Yes.
21   Q.   So looking at this proposal, there
22 are only four datasets who have epochs -- strike

Page 260

1 that.
2   So what is -- if we take a look at
3 the first row under epochs for Stack Exchange, it
4 says 2.903 and then a bunch of numbers, right?
5   A.   Yes.
6   Q.   Can you explain to us what that
7 means?
8   A.   I didn't precisely implement this.
9 I only have a high level understanding what of
10 what it means.
20   Q.   When you -- maybe -- I think this is
21 my fault. Let me back up a bit.
22   When you said relative proportion of

65 (Pages 257 to 260)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 261

1   that dataset in the model, did you mean a higher
2   relative proportion of that dataset vis-à-vis the
3   entirety of the data pre-training corpus?
4   A.   I forget precisely what I said, but
5   increasing the epochs would have the effect of
6   increasing the proportion of tokens seen from
7   that dataset.
8   Q.   So looking at this column of epochs
9   here, right, ███████████████████████████████
    ███
    ███
    ███████████████████████████████████████
13   A.   Correct.
    ███
    ███
    ███
    ████████████████████████
18   A.   Not necessarily, because some of
19 these datasets are smaller than others.
20   Q.   So if a dataset is smaller, would
21 increasing the number of epochs sort of balance
22 out the dataset vis-à-vis larger datasets in the

Page 262

1   corpus?
2   A.   Can you be more precise, what you
3   mean balance out.
4   Q.   I think I will move on. Can you go
5   back to the document marked as Plaintiffs' 16,
6   please?
7   A.   Yes.
8   Q.   So then you say -- you respond -- so
9   do you see the message there, share the document
10 we were just examining was marked as
11 Plaintiffs' 17, right? Guillaume's 7:16:34
12 message?
13   A.   Yes, I see the message.
14   Q.   You respond to Guillaume, right?
15 You say: The sheet is slightly confusing because
16 for example -- e.g. -- for example, right, I see
17 Books3, Gutenberg listed both as B2 run and won't
18 use.
19   Right?
20   Do you see that?
21   A.   I see that message, the document.
22   Q.   Then you share a PNG, right, at

Page 263

1   7:16:43 PDT, right?
2   A.   It looks like it, yes.
3   MR. YOUNG: So I am going to mark as
4   Plaintiffs' 18 --
5   (Whereupon, Kambadur Exhibit 18, PNG
6   file Bates Number Meta Kadrey 00074157, was
7   marked for identification.)
8   MR. YOUNG: Document bearing Bates
9   label Meta Kadrey 00074157 as 18A, metadata
10 associated with that document.
11   (Whereupon, Kambadur Exhibit 18A,
12 Metadata associated with Exhibit 18, was
13 marked for identification.)
14   MR. YOUNG: I think we are closing
15 on an hour. I will take one more break and
16 then I think we can close this out.
17   MS. STAMESHKIN: Great.
18   MR. WEINSTEIN: Sounds good.
19 BY MR. YOUNG:
20   Q.   So, Ms. Kambadur, sorry for that.
21 Let me when you're -- you've had a chance to
22 review the document and you're ready to proceed.

Page 264

1   A.   Yes, I've reviewed the document, at
2   least scanned it.
3   Q.   Okay. Do you recognize this
4   document?
5   A.   Yes, at least the document that it's
6   a screenshot of.
7   Q.   Okay. And is this document the
8   screenshot of the same document that Guillaume
9   previously shared as -- that we were looking at
10 as Plaintiffs' 17?
11   A.   Yes, it looks like.
12   Q.   Okay. And is this the same document
13 that you shared in the post that we were
14 referencing in Plaintiffs' 16, the message
15 thread?
16   A.   I'm not sure.
17   Q.   Okay. Will you please look at
18 what's been marked as Plaintiffs' 18 and if you
19 can look at the row labeled "File name" and see
20 that there's a PNG file, correct?
21   A.   Yes.
22   Q.   Is that the same PNG file name as

66 (Pages 261 to 264)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 265

1 your message that you sent to Guillaume Lample at
2 7:16:43 PDT in Plaintiffs' 16?
3   A.   It appears to be, yes.
4   Q.   Does that suggest to you that
5 Plaintiffs' 18 is the same document that you
6 shared with Guillaume Lample at 7:16:43 PDT?
7   A.   That's what I would assume.
8   Q.   Okay. So before we move back to
9 Plaintiffs' 18, do you see your message at
10 7:16:42, where you say: The sheet is slightly
11 confusing?
12   A.   Yes, I see the message.
13   Q.   Okay. And then you explain why it's
14 confusing, right?
15   A.   Yes.
        ██████████████████████████████
        ███████████ 2 Rn and won't use, right?
18   A.   Yes.
19   Q.   Okay. And if we go back to
20 Plaintiffs' 18, we see Rows 35, 36 and 37 in
21 Plaintiffs' 18.
22   A.   Yes.

Page 266

1   Q.   Okay. Those rows were cut off from
2 Guillaume's screenshot at Plaintiffs' 17, weren't
3 they?
4   A.   Yes.
   ██
   ██
   ██
   ████████████████████████████████
   ██
   ████████████
   ██
   ██
   ████████████████████████████████
   ██
   ██
   ████████████████████████████
   ██
   ██
   ████████

Page 267

1 ██
2 ██
3 █████████████.
4   Q.   Okay. So Row 37, do you see where
5 it says Gutenberg?
6   A.   Yes.
7   Q.   Do you have an understanding of what
8 that is, Gutenberg?
9   A.   My understanding is it would have
10 some relation to the Gutenberg books dataset.
11   Q.   It's Project Gutenberg that we were
12 referring to earlier?
13   A.   Yes.
   ████████████████████████████████
   ██
   ██
   ████████████████████████████████
   ████████████████████████████████████
   ████████████████████████████████
   ████████████████████████████████████
21   Q.   Yes. Yes.
22   Were you confused because you

Page 268

1 ██
   ████████████████
3   A.   I don't remember why I was confused.
4   Q.   But Guillaume offers you an
5 explanation, right, at 7:17:33?
6   A.   Yes.
7   Q.   Okay. And Guillaume says:
   ██
   ████████████████████████████████
10   Right?
11   A.   That's what he said.
   ██
   ██
   ████████████████████████
   ██
   ████████████████████████████████████
   ██
   ████████████████
20 BY MR. YOUNG:
21   Q.   Okay. Then what do you understand
22 it to mean?

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 269

1   A. That it's some combination of some
2 versions of those datasets with some application
3 of de-duplication.

[REDACTED]

Page 270

1 BY MR. YOUNG:

[REDACTED]

11   Q. So when you say de-duplication,
12 you're still leaving at least one version of that
13 work in there, right?
14   A. At least one version of what work.
15   Q. So let me -- let me -- let me strike
16 that and let me try that again.
17   So we talked about de-duplication
18 earlier, right?
19   A. Yes.
20   Q. Okay. Okay. So when you're talking
21 about de-duplication, you're never removing all
22 of the things that you're trying to de-duplicate,

Page 271

1 right?
2   A. Generally, when we talk about
3 de-duplication on a document level, we're
4 retaining one version of a document.

[REDACTED]

16   MR. YOUNG: Okay. I think we can
17 take a break now.
18   THE VIDEOGRAPHER: The time right
19 now is 4:03 p.m. We are off the record.
20   (Whereupon, a brief recess was
21 taken.)
22   THE VIDEOGRAPHER: The time right

Page 272

1 now is 4:19 p.m. We are back on the record.
2   MR. YOUNG: We will mark as
3 Plaintiffs' 19, a document bearing Bates
4 label Meta Kadrey 00063689.
5   (Whereupon, Kambadur Exhibit 19,
6 E-mail thread Bates Number Meta Kadrey
7 00063689, was marked for identification.)
8 BY MR. YOUNG:
9   Q. Ms. Kambadur, please take a moment
10 to review the document, and let me know when
11 you're ready.
12   A. I've reviewed the document.
13   Q. Thanks.
14   Ms. Kambadur, do you recognize this
15 document?
16   A. I have a vague recollection.
17   Q. Is this document an e-mail thread
18 that you participated in on or about
19 December 23rd, 2022?
20   A. It looks like I was at least CC'd on
21 this e-mail.
22   Q. And you were -- -- if you look at

68 (Pages 269 to 272)

Page 273

1   the top e-mail, you were also added, right?

2   A.   Yes.

3   Q.   Do you understand that someone

4   wanted to get your attention on this e-mail?

5   A.   I guess that's the intent.

6   Q.   Do you understand these e-mails to

7   have been sent in the regular course of business?

8   A.   Yes.

9   Q.   And what is the subject of this

10 e-mail thread?

11   A.   Do you want me to read it?

12   Q.   Yes.

Page 274

1   Q.   Can you turn to the bottommost

2   e-mail, the first in time e-mail.

3   Is that the e-mail sent by

4   Mr. Rodriguez on December 16, 2022, at time stamp

5   12:54 a.m.?

6   A.   Yes, I see that e-mail.

7   Q.   Could us remind me who

8   Aurelien Rodriguez is?

9   A.   He was a research manager at Meta.

10   Q.   So a research manager. So as of

11 this e-mail, was he equal in seniority to you?

12   A.   I am trying to -- I think I may have

13 been a level higher than him, but we were equal

14 in the reporting chain.

15   Q.   Okay. Now, Aurelien begins his

16 e-mail -- Mr. Rodriguez begins his e-mail. It

17 says:

19   What does -- what does ablation mean

20 in a generative AI context?

21   A.   In this context, it means a

22 controlled experiment.

Page 275

1   Q.   So what -- a controlled experiment

2   how?

3   A.   Sorry. I don't understand the

4   question.

5   Q.   Can -- can ablation mean including

6   certain data?

7   A.   An ablation is more just a general

8   term for a controlled experiment. So you could

9   change the data as one of the variables that you

10 choose to change.

11   Q.   So, for example, which datasets are

12 included or are not included in the experiment?

13   A.   Yes, that could be an example

14 experiment.

15   Q.   What about like weighting a certain

16 dataset?

17   A.   That could also be an example

18 experiment, yes.

19   Q.   Okay. So based on this first

20 sentence,

Page 276

20   Q.   What's the date on this e-mail?

21   A.   December 16th, 2022.

22   Q.   So this was before the release of

Page 277

1   the Llama 2 models, correct?
2   A.   Correct.
3   Q.   Now, do you see how Aurelien signs
4   off his e-mail.
5   It says: Cheers, the Genesis team.
6   Right?
7   A.   Yes.
8   Q.   And do you recall earlier telling us
9   that "Genesis" was another name for the Llama
10   models?
11   A.   In this context, it's -- they're
12   also referring to it as a team, but it has -- we
13   had internal versions of the model named Genesis
14   that were the initial basis for Llama models.
15   Q.   So Aurelien is reporting on the
16   results or his interpretation of the experiments
17   on behalf of Genesis team, right?
18   A.   That's what it appears in this
19   e-mail.
20   Q.   Do you know if the Genesis team
21   eventually became the Llama team?
22   A.   Not precisely. Well, some members

Page 278

1   of the Genesis team joined -- or what I think he
2   is referring as the Genesis team. I am not sure
3   this was, like, an official organization. But
4   some members of the group I think he is referring
5   to joined a larger group which then worked on
6   Llama.
7   Q.   So do you see the sentence in bold?
8   He says: We are confident that we can reproduce
9   Chinchilla with OSS data.
10   Do you see that sentence -- or that
11   clause? Excuse me?
12   A.   Yes.
13   Q.   Is Chinchilla a model that was
14   produced by Google?
15   A.   Yes.
16   Q.   So Aurelien is saying that -- based
17   on this sentence here, do you understand Aurelien
18   to be attempting or the Genesis team to be
19   attempting to reproduce Chinchilla?
20   MR. WEINSTEIN: Object to form.
21   A.   Yes, I understand that this was a
22   goal of the Genesis model to reproduce some

Page 279

1   metrics that were similar to Chinchilla.
2   BY MR. YOUNG:
3   Q.   So when you -- what do you
4   understand the word "reproduce" to mean in this
5   context?
6   A.   Based on a broader context I have
7   from this e-mail, I understand it to mean using
8   relatively similar model size scales to achieve
9   similar benchmark performance.
10   Q.   Do you know what OSS data means?
11   A.   We use OSS sometimes as a shorthand
12   for open source, so that would be my presumption,
13   but I don't know what precisely Aurelien means.
14   Q.   Is open source data synonymous with
15   publicly available data?
16   MR. WEINSTEIN: Object to form.
17   A.   I don't know what Aurelien means in
18   his reference.
19   BY MR. YOUNG:
20   Q.   Do you have an understanding of what
21   open source data is?
22   A.   A high level understanding.

Page 280

1   Q.   What is your understanding of what
2   open source data is?
3   A.   Data which has been permissibly
4   licensed.
5   Q.   Can you explain what "permissibly
6   licensed" means?
7   A.   I don't have a precise definition,
8   but which has a license attached that gives
9   others permission to use that artifact.
10   Q.   So Aurelien says: And we will train
11   our model only on data compatible with open
12   source.
13   Do you see that?
14   A.   Yes, I see that.
15   Q.   Do you see the next sentence? He
16   goes: We are excited that this allows us to
17   align Genesis current milestone of reproducing
18   Chinchilla with our FRESCO value.
19   Do you know what FRESCO value means?
20   A.   I believe FRESCO was an acronym used
21   for some kind of mission or mission statement or
22   something like that in FAIR.

70 (Pages 277 to 280)

Page 281

1   Q.   Do you have a recollection of what
2 that acronym stood for?
3   A.   I don't recall for sure.
4   Q.   Do you recall what some of the
5 letters stood for?
6   A.   I believe O stands for open. That's
7 the only one I'm remembering off the top of my
8 head, and I am not even 100 percent sure on that
9 one.
10   Q.   Okay. Do you see how further on in
11 that paragraph Aurelien talks about RAI?
12   A.   Yes.
13   Q.   Is RAI Responsible AI?
14   A.   I believe that he is using it as a
15 synonym -- as a abbreviation for Responsible AI,
16 yes.
17   Q.   Is that a division at Meta?
18   A.   At some point, we had an
19 organization called RAI. I am not sure if that's
20 what Meta -- or if that is what is Aurelien is
21 referring to here.
22   Q.   Do you know if that organization

Page 282

1 still exists?
2   A.   There are still members of that
3 organization, but there have been an
4 organizational changes since then.
5   Q.   Do you know who was in the RAI
6 organization as of this e-mail?
7   A.   No.
8   Q.   Do you know who is in the RAI
9 organization now?
10   A.   The -- the -- there is not to my
11 knowledge an organization called RAI.
12   Q.   Is there an organization -- so what
13 do you understand the role of the Responsible AI
14 organization to have been?
15   A.   I don't have that broad of context
16 to understand their complete role at Meta.
17   Q.   So then Aurelien continues. He
18 makes at -- Luke Zettlemoyer. Do you know who
19 Luke Zettlemoyer is?
20   A.   Yes.
21   Q.   Who is Luke Zettlemoyer?
22   A.   He's a research director at FAIR.

Page 283

1   Q.   Is he a full-time employee of Meta?
2   A.   I am not sure if he is full-time or
3 not. He is co-listed as a professor.
4   Q.   So it goes on: Your experience with
5 open source at OPT will be greatly appreciated
6 here.
7   What is OPT?
8   A.   OPT is another model that we trained
9 on Meta.
10   Q.   What type of model was that?
11   A.   Also a language large model.
12   Q.   Was that model ever released to the
13 public?
14   A.   Yes.
15   Q.   What was the name?
16   A.   OPT.
17   Q.   Thank you for that. I realized that
18 was a silly question after I asked it.
19   So if you go down to -- after -- do
20 you see the line after the Genesis team, the
21 sign-off?
22   A.   Yes.

Page 284

1   Q.   It says: Details of the experiment.
2   Do you see that section?
3   A.   I see that section.
4   Q.   Do you see the reference to the XL
5 Former's team?
6   A.   Yes.
7   Q.   Do you know what that is?
8   A.   I know that XL Former is the name of
9 a library. I am not sure who he is referring to
10 when he says XL Former's team.
11   Q.   Now, before we go into this, can we
12 go back to the final paragraph right before the
13 cheers, Aurelien signed off.
14   So the first sentence of that final
15 paragraph, do you see that sentence begins: Once
16 our final training is done?
17   A.   Yes.
18   Q.   So he says: ██████████
   ████████████████████████████
   ███████████████████████████
   █████████████████
22   Do you see that?

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 285

1   A.   Yes.
2   Q.   Do you understand him to understand
3   license data to be different than open source
4   data?
5   A.   From that sentence, it's hard to
6   tell what his understanding was.
7   Q.   So that first sentence underneath
8   details of experiment,

11   Do you see that?
12   A.   Yes, I see that.

21   Q.   And do you understand that to be
22 distinct from the OSS data he was referring to

Page 286

1   earlier?
2   A.   It appears to be distinct based on
3   his text, but from the text, my understanding is
4   he also doesn't have a precise understanding of
5   the meaning of those terms.
6   Q.   When he is -- based on the context
7   of this e-mail, it seems pretty clear that when
8   he's talking about open source data, he is
9   referring to

11   A.   Perhaps, based on the context of
12 this e-mail, but there -- I have an additional --
13 yeah, can you rephrase that little bit because I
14 am -- actually I'm unclear.
15   Q.   Aurelien in that bold statement
16 says: We can reproduce Chinchilla with OSS data
17 and we will train our models only on dates
18 compatible with open source.
19   Is that correct?
20   A.   Yes, he says that.
21   Q.   So if he's -- there's an
22 understanding he has of what OSS data is,

Page 287

1   correct?
2   A.   He has made some definition of open
3   source data.
4   Q.   And then if we look at the paragraph
5   at bottom of the page where he is describing the
6   experiment, he's saying that the --

Page 288

5   MR. WEINSTEIN: Object to form.
6   A.   I don't really understand what he
7   meant partially because even open source data, to
8   my understanding, has licenses or can a have a
9   license.
10 BY MR. YOUNG:
11   Q.   Can you sometime buy licenses to use
12 data?
13   MR. WEINSTEIN: Object to form.
14   A.   The wording is a little confusing to
15 me. Like, it can -- I am not sure exactly how
16 the -- how the legal purchasing of the license
17 worked, but I know that you can purchase data.
18 BY MR. YOUNG:
19   Q.   Can you purchase licenses to, for
20 example, use certain media?
21   MR. WEINSTEIN: Object to form.
22   A.   Can someone purchase a license to

72 (Pages 285 to 288)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 289

1 utilize media? I think so, but I am not an
2 expert.
3 BY MR. YOUNG:
4   Q.   For example, have you ever -- rent a
5 streaming video?
6   A.   Yes, I have rented streaming videos.
7   Q.   Do you understand that to be
8 purchasing a license to that video?
9   MR. WEINSTEIN: Object to form.
10   A.   I have never thought about it that
11 way. I don't -- I have never seen a specific
12 license attached.
13 BY MR. YOUNG:
14   Q.   Okay. So continuing on to the next
15 page, so Aurelien continues to describe this
16 experiment.
17   You -- so what is ███████████
██████████████████████████████████████
██████████████████████████████████████
██████████████████████████████████████
██████████████████████
22   Q.   And there's a number of -- there's a

Page 290

1 chart describing how the model performed under
2 each of those benchmarks, correct?
3   A.   How this -- how the experiment --
4 experimental model performed, yes.
5   Q.   And there generally are higher
██
██████████████████████████████████████
██
██████████████████████████████████████
██████████████████████████████████
██
██████████████
██████████████████████████████
██
██████████████████████████████████████
██████████████████
██████████████████████████████████████
████████████
20   Q.   Okay. So before we examined that,
21 he is -- Aurelien is discussing -- it says that
22 these are model versions of Genesis, right?

Page 291

1   Do you see at that? Right -- right
2 before the table?
3   A.   Yes, I see that.
4   Q.   Do you understand that to also mean
5 model versions of what become Llama?
6   A.   I understand that to mean similar
7 architectural settings as Llama, as what would
8 become the future Llama 1.
9   Q.   Okay. So each of the rows describe
10 data that was either included or not included in
11 each of the experiments. Would that be fair?
12   A.   That's what it looks like.
██████████████████████████████████████
██████████████████████████████████████
██████████████
██
██
██████████████████████████████████████
████████████████████████
██████████████████████████████
██

Page 292

██████████
██████████████████████
██████████████████████████████████
██████████████████████████
██
████████████████████████
██
██████████████████████████
██████████████████████████████
██████████████████████████████████████
████████████████████████████████
██████████
██
████████████
██████████████████████████████████
██
████████████

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 293

Page 295

13   Q.   So can you turn back to the first
14 page, the very bottom e-mail from Joelle Pineau.
15   Do you know who Joelle Pineau is?
16   A.   Yes.
17   Q.   Who is she?
18   A.   I don't know her title at the time
19 of this e-mail, but she is currently VP of FAIR.
20   Q.   Then she sends an e-mail on
21 December 16th, 2022 at 8:43 a.m., right?
22   A.   Yes.

Page 294

Page 296

1   Q.   She adds you, right, to the e-mail
2   thread?
3   A.   Yes.
4   Q.   And Mary, right?
5   A.   Yes.
6   Q.   And Mary is Mary Williamson?
7   A.   It appears so.
8   Q.   Is Mary Williamson a member of FAIR
9   at this time?
10   A.   Yes.
11   Q.   So Joelle Pineau says: Thanks for
12 the meeting. Great to have this analysis help
13 guide our investments.
14   Do you see that?
15   A.   Yes.
16   Q.   Do you have an understanding of what
17 Ms. Pineau meant by "investments"?
18   A.   I couldn't be certain. It could
19 mean the model training investments.

am

74 (Pages 293 to 296)

Page 297

1   not certain.

2   Q.   And for pre-training models?

3   A.   It's hard to speculate from the

4   limited context here.

5   Q.   Okay.

6   A.   It could also mean whether we should

7   open source this model. There's a lot of

8   interpretations here.

9   Q.   You can set that document aside.

10   MR. YOUNG: So I am going to mark as

11   Plaintiffs' 19 -- 20. Apologize -- a

12   document bearing Bates label Meta Kadrey

13   0054898.

14   (Whereupon, Kambadur Exhibit 20,

15   Workplace Chat Bates Number Meta Kadrey

16   0054898, was marked for identification.)

17   BY MR. YOUNG:

18   Q.   Why don't you take a moment to

19   review the document and let me know when you are

20   ready.

21   A.   I reviewed it.

22   Q.   Great.

Page 298

1   Does Plaintiffs' -- Plaintiffs' 20

2   appear to be a Workplace Chat between you and

3   Nisha Dayal taking place on or around

4   August 18th, 2023?

5   A.   Yes.

6   Q.   Were these chats generated in the

7   regular course of business?

8   A.   Yes.

9   Q.   Who is Ms. Dayal? Is that how you

10   pronounce her name?

11   A.   I am not sure.

12   Q.   I am going to call her Nisha.

13   Do you know who Nisha is?

14   A.   Yes.

15   Q.   What is Nisha's role at Meta?

16   A.   I don't know her specific title, but

17   she is on our communications team.

18   Q.   And Nisha is still with Meta; is

19   that correct?

20   A.   I am not sure.

Page 299

2   Right? Do you see that?

3   A.   Yes.

5   Right?

6   A.   Yes.

9   Q.   Now, do you see Nisha's message at

10   13:41:11 PDT a couple of messages down?

11   A.   Yes.

12   Q.   She says: Questions they posed

13   include --

14   Sorry, let's back up.

15   Her message from 13:44:01, right?

16   A.   Uh-huh.

17   Q.   So Nisha is saying that she received

18   an inquiry The Atlantic.

19   Do you see that?

20   A.   Yes, I see that.

21   Q.   Do you understand The Atlantic to be

22   the magazine The Atlantic?

Page 300

1   A.   I don't know what kind of

2   publication it is.

3   Q.   Journalists?

4   A.   I knew it was some kind of article.

5   Q.   And they asked about the use of

6   Books3 in Llama, right?

7   A.   According to Nisha, yes.

8   Q.   So Nisha is asking you to help with

9   responding The Atlantic, it seems like, right?

10   A.   I think she's asking me for more

11   technical details so that she can prepare a

12   response.

13   Q.   So going back to her message bearing

14   time stamp 13:44:11 PDT, now, she relays to you a

15   number of questions posed by The Atlantic,

16   correct?

17   A.   Yes.

18   Q.   So the very first question is: The

19   research paper about the construction of Llama

20   clearly states that Books3 is among the corpora

21   of the program's training data.

22   Right?

75 (Pages 297 to 300)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 301

1   A.   That sentence is there. It's not a
2   question.
3   Q.   Right. Sorry.
4   Do you know which research paper she
5   is referring to?
6   A.   I don't know think -- I don't know
7   who is referring to which Llama in this sentence.
8   It looks like a quote from Nisha.
9   Q.   Okay. Do you recall if the Llama 1
10 paper stated the Books3 was amongst its
11 pre-training corpora?
12   A.   I would have to double-check our
13 papers or the paper.
14   Q.   Do you still have Plaintiffs' 1 in
15 front of you, the Llama 1 paper?
16   So on page -- do you see the page
17 bearing Bates label Meta Kadrey 000079?
18   A.   Yes.
19   Q.   Do you see the Section 2.1
20 pre-training data?
21   A.   Yes.
22   Q.   Do you see a section titled

Page 302

1   "Gutenberg and Books3"?
2   A.   Yes.
3   Q.   Does that suggest to you that the
4   Llama 1 paper is saying that Books3 comprises the
5   pre-training data for Llama 1?
6   A.   The Books3 section of the Pile, it's
7   saying is in the pre-training data.
8   Q.   Do you think that's the -- that's
9   the research paper that Nisha referring to in
10 Plaintiffs' 20?
11   A.   Again, I think Nisha is quoting a
12 journalist here. So I don't know if Nisha
13 herself is referring to anything.
14   Q.   Okay. But do you think that's --
15 that's the section of the research paper that the
16 journalist was referring to?
17   A.   It's hard to know for sure, but that
18 would be a reasonable assumption.
19   Q.   Okay. Now, Nisha also asks later at
20 13:47 PDT to have someone named Ahuva, quote,
21 looped in.
22   Do you see that?

Page 303

1   A.   Yes, I see that.
2   Q.   Who is Ahuva?
3   A.   Ahuva is one of our legal counsel at
4   Meta.
5   Q.   Okay. She describes an escalation.
6   What is an escalation?
7   A.   I don't know precisely what she is
8   referring to, but it may mean informing her
9   manager.
10   Q.   And Nisha also says she is trying to
11 understand, quote/unquote, the ground truth,
12 right?
13   A.   Yes, she said that.
14   Q.   Now, you had also used that phrase
15 "ground truth" earlier today.
16   Do you recall?
17   A.   Yes, I recall.
18   Q.   Can you remind us what ground truth
19 means?
20   A.   I think these are slightly different
21 contexts. So do you want a broader definition,
22 or do you want me to go back to the context we

Page 304

1   were discussing earlier or discuss this context?
2   Q.   Let me try again. What do you
3   understand Nisha to mean when she says "ground
4   truth"?
5   A.   I think she wants to -- it's hard to
6   precisely know what she means, but my
7   understanding from reading this now is that she
8   is trying to gather additional facts from our
9   research team.
10   Q.   In order to relay to the journalist
11 of The Atlantic?
12   A.   I am not sure what she ultimately
13 relayed or how she would choose what to relay.
14   Q.   Towards the bottom of the page and
15 it goes on to the next page, Nisha shares a GIF
16 file.
17   Do you see that?
18   A.   Yes.
19   MR. YOUNG: Okay. I am going to
20 mark Plaintiffs' 21 --
21   (Whereupon, Kambadur Exhibit 21,
22 Screenshot Bates Number Meta Kadrey 00054019,

76 (Pages 301 to 304)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 305

1   was marked for identification.)
2   MR. YOUNG: -- what was produced
3   natively as Meta file 000024898. This is, I
4   believe, a screenshot. Oh, I am sorry. It's
5   Meta Kadrey 00054019. That's Plaintiffs' 21.
6   I misspoke. And 21A is metadata associated
7   with that document, Plaintiffs' 21.
8   (Whereupon, Kambadur Exhibit 21A,
9   Metadata associated with Exhibit 21, was
10   marked for identification.)
11   BY MR. YOUNG:
12   Q.   Do you recognize this document?
13   A.   I don't recall it.
14   Q.   Okay. Take a look at 19A.
15   A.   19A?
16   Q.   Yep, marked as 19A, it should be the
17   metadata associated with that file.
18   A.   Are we talking about 21A?
19   MR. WEINSTEIN: 21A.
20   BY MR. YOUNG:
21   Q.   I am sorry. 21A. It's been a long
22   day. I apologize.

Page 306

1   And you see the file name of 21A,
2   right?
3   A.   Yes.
4   Q.   Okay. It's a GIF -- the name of a
5   GIF file, right?
6   A.   It looks like it.
7   Q.   And if you go back to Plaintiffs'
8   21, is the GIF file name identical to the GIF
9   that Nisha shared with you?
10   A.   It appears to be.
11   Q.   Does that suggest to you that the
12   GIF marked as Plaintiffs' 21 is the same as the
13   one Nisha sent you in Plaintiffs' -- Plaintiffs'
14   20?
15   A.   That would be my assumption.
16   Q.   Do you recognize this GIF?
17   A.   I don't recall it. I think I have
18   seen that GIF in general, but I don't recall it
19   in this conversation context.
20   Q.   Do you understand this to be an
21   image of Kate McKinnon impersonating Hillary
22   Clinton?

Page 307

1   A.   That's what it looks like.
2   Q.   The text says: And that is how you
3   pivot.
4   Right?
5   A.   Yes, pivot, yes.
6   Q.   How you pivot, right.
7   Do you have an understanding of what
8   "pivoting" means in this context?
9   A.   In the context of the GIF or the
10   context of the meme or the context from Nisha?
11   Q.   The GIF or meme.
12   A.   I don't know for sure. I don't
13   think I have seen this skit.
14   Q.   Do you understand -- you described
15   this as a meme, right?
16   A.   Yes.
17   Q.   So do you understand this meme to be
18   poking fun at Hillary Clinton who was at that
19   time dodging or avoiding questions in debates?
20   A.   I don't know. I don't watch many
21   presidential debates.
22   Q.   Do you understand the meaning of --

Page 308

1   the context of this meme?
2   A.   I have a guess that it may be on
3   Saturday Night Live because it's Kate McKinnon
4   and it says NBC.
5   Q.   Do you understand that pivoting
6   means to avoid or dodge questions?
7   A.   I don't know -- I don't think that's
8   how I would define pivot.
9   Q.   Did you find this meme amusing when
10   Nisha sent it to you?
11   A.   I don't recall.
12   Q.   Did you recall having -- recall
13   having an understanding of what Nisha was trying
14   to communicate to you when she sent you this
15   meme?
16   A.   I don't recall, but it seems
17   unrelated to some of other conversation based on
18   her next sentence.
19   Q.   Now, if we go a few messages up, do
20   you see your message at 13:51:10 PDT?
21   A.   Yes.
22   Q.   You say: Yes, I went to an AI event

77 (Pages 305 to 308)

Page 309

1 yesterday with a bunch of people from other
2 companies, and any time I mentioned I work on
3 Llama 2, they kept asking me about datasets,
4 smiley face. I was just like see the paper/no
5 comment, LOL.
6   Right?
7   A.   Yes.
8   Q.   Do you think that might be what
9 Nisha was referring to?
10   A.   It's really hard to say because
11 there are additional redacted comments that I
12 don't recall.
13   Q.   So going back to that second page of
14 that document, now Nisha starts talking --
15 returns to the subject of Books3, right?
16   A.   Yes.
17   Q.   She asks if -- do you know if any of
18 -- if assurances from them that their catalog
19 didn't contain any pirated works.
20   Right?
21   A.   That's what she typed.
22   Q.   Do you have an understanding of what

Page 310

1 pirated work means?
2   A.   I don't know for sure what it means
3 in this context.
4   Q.   Does it mean copyrighted works taken
5 without permission?
6   MR. WEINSTEIN: Object to form.
7   A.   I don't know how you can draw that
8 conclusion from this information.
9 BY MR. YOUNG:
10   Q.   Do you have an understanding of what
11 the terms "pirated works" mean?
12   A.   I don't know what this particular
13 journalist that Nisha is quoting meant by
14 "pirated works," no.
15   Q.   Did you have an understanding -- you
16 yourself have an understanding what "pirated
17 works" mean?
18   A.   It's a very colloquial term. So I'm
19 not exactly sure what the precise definition is.
20   Q.   Would it mean, for example, perhaps
21 taken without the copyright holder's permission?
22   A.   I suppose someone could -- could

Page 311

1 define it as that.
2   Q.   Now, moving on to your statement as
3 of 14:08:11.
4   Do you see that, that statement
5 there "no worries"?
6   A.   Yes.
7   Q.   You say: No worries.
8   You mean for distributing Llama 1,
9 right?
10   A.   Yes.
11   Q.   Okay. And this was in response to
12 Nisha's questions two questions above, "Do we
13 have that same license on Llama 1," right?
14   A.   Yes.
15   Q.   Okay. What did you mean by
16 "distributing Llama 1"?
17   A.   I think I was trying to distinguish
18 what license she was referring to and whether
19 that was the license attached to Llama itself.
20   Q.   By "distributing Llama 1," did you
21 mean -- mean make available for use?
22   A.   I suppose so.

Page 312

1   Q.   And then you go -- you respond "no,"
2 right, because Llama 1 was a research-only
3 model -- model, right? And Llama 2 is
4 commercial, right?
5   Do you see that?
6   A.   Yes.
7   Q.   So what you're saying is that you --
8 Llama 1 and Llama 2 were being distributed under
9 different licenses, right?
10   A.   That's what I'm saying, yes.
11   Q.   And going back further, right,
12 you're -- the start of this conversation was a
13 discussion about distribution of the Books3
14 dataset, right?
15   Do you see Nisha -- Nisha's message
16 of 14:04:16?
17   A.   Sorry. Could you repeat that
18 question?
19   Q.   Yeah. So let's actually go to your
20 message of 14:05:13.
21   Do you see that message?
22   A.   Yes.

Page 313

1   Q.   Okay. And you're responding to
2 Nisha, right, who asks about assurances about
3 Books3 dataset, right, at 14:04:16; is that
4 right?
5   A.   Yes.
6   Q.   Okay. And then you provide a URL,
7 right, which appears to be a hyperlink to a
8 license, right?
9   A.   Yes.
10   Q.   And that appears to be a license
11 from a EleutherAI, right?
12   A.   Yes.
13   Q.   Because GitHub.com/EleutherAI,
14 right, that's how you can tell?
15   A.   Yes, that would be my assumption.
16   Q.   Okay. And then it's, thought goes
17 on per Pile, right?
18   A.   Yes.
19   Q.   So does that tell you that that's
20 the license for use of the Pile?
21   A.   That would make sense. I -- I can't
22 be certain.

Page 314

1   Q.   Do you know if EleutherAI sought
2 permission from anyone to include their works in
3 the Pile?
4   MR. WEINSTEIN: Object to form.
5   A.   I wasn't involved with EleutherAI,
6 so I -- I don't know how they collected data.
7   Q.   Do you know if they paid anyone any
8 money to include any works in the Pile?
9   A.   Again, I -- I wasn't involved with
10 EleutherAI or their collection, so I do not know.
11   Q.   Okay. Now, your -- do you see
12 Nisha's message at 14:20:12?
13   A.   Yes.
14   Q.   And she goes: Otherwise, the fact
15 that Books3 allegedly has issues with copyrighted
16 works is not new.
17   Right?
18   Do you see that?
19   A.   Yes.
20   Q.   Do you agree with her statement as
21 of August 18, 2023, that Books3 had issues with
22 copyright as of that point?

Page 315

1   MR. WEINSTEIN: Object to form.
2   A.   Sorry. Your -- your question is not
3 matching what Nisha said, so I don't know how to
4 answer that.
5 BY MR. YOUNG:
6   Q.   Okay. Well, you respond to Nisha,
7 right, at 14:24:13?
8   A.   Yes.
9   Q.   You say: I feel like that's on the
10 people who distributed the dataset somewhat, no?
11   Right?
12   A.   Yes.
13   Q.   So would -- were you laying blame on
14 the EleutherAI if -- if the Pile contained
15 copyrighted work?
16   A.   I don't know that I would -- yeah.
17 I don't know that I can make that statement. I
18 don't recall exactly what I meant.
19   Q.   So the "people" -- the "people" in
20 the sentence, do you think that refers to
21 EleutherAI?
22   A.   I suppose.

Page 316

1   Q.   And then 14:31:58, do you see that
2 sentence there beginning with "um"?
3   A.   Yes.
4   Q.   Okay. And then you say: ████
  ██ ████████████████████████████
  ██ ██████████████████████████
  ██ ████████████████
8   Do you see that?
9   A.   Yes.
10   Q.   Okay. Do you recall us test --
11 discussing this very early in this deposition
12 about removing kind of the front pages of the
13 books?
14   A.   I recall discussing this area.
15 ██████████████████████
  ██ ████████████████████
  ██ ██████████████████
  ██ ████████████████
  ██ ██████████████████
  ██ ██████████████████
  ██ ██████████████████████
  ██ ██

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 317

1   MR. YOUNG: Okay. I'm done with
2   this document and I've got one more line of
3   questioning. So about -- maybe about an half
4   hour left. So I'll leave it to you guys.
5   Do you guys want to take a short
6   break or do you guys want to keep trekking?
7   THE WITNESS: I'm either way.
8   MR. WEINSTEIN: We can proceed, I
9   think.
10   MR. YOUNG: Okay. Great.
11   And then we are going to mark as
12   Plaintiffs' 22, a document bearing Bates
13   label Meta Kadrey 00054518.
14   (Whereupon, Plaintiffs' Exhibit 22,
15   Bates Number Meta Kadrey 00054518, was marked
16   for identification.)
17   BY MR. YOUNG:
18   Q.   So take a couple moments to review
19   the document and let me know when you're ready.
20   A.   I reviewed the document.
21   Q.   All right. And is Plaintiffs' 22 a
22   Workplace Chat between you and yourself -- you

Page 318

1   and others on or around June 2, 2023?
2   A.   That's what this looks like, yes.
3   Q.   And are these message sent in the
4   ordinary course of business?
5   A.   Yes.
6   Q.   And the very first message was from
7   you, right?
8   A.   In this exhibit, yes.
9   Q.   And it say -- you say: Any chance
10   we can fix this behavior? Looks pretty bad, even
11   if it's hallucinating.
12   Right?
13   A.   Yes.
14   Q.   Do you recall what behavior you were
15   trying to fix?
16   A.   Not specifically.
17   Q.   Okay. And then you share three
18   files at 11:44:20.
19   Do you see that?
20   A.   Yes.
21   MR. YOUNG: So I am going to mark as
22   Plaintiffs' 23, a document bearing Bates

Page 319

1   label Meta Kadrey 00054520.
2   (Whereupon, Kamadur Exhibit 23,
3   Bates Number Meta Kadrey 00054520, was marked
4   for identification.)
5   MR. YOUNG: And then, Madam Court
6   Reporter, will you please mark as 23A, the
7   Metadata associated with this document.
8   (Whereupon, Kamadur Exhibit 23A,
9   Metadata associated with Exhibit 23, was
10   marked for identification.)
11   BY MR. YOUNG:
12   Q.   Let me know when you've reviewed the
13   document, Ms. Kambadur.
14   A.   Yes, I have reviewed the document.
15   Q.   Do you recognize this document?
16   A.   Yes.
17   Q.   What is it?
18   A.   It appears to be a screenshot of
19   model user prompts and model outputs from one of
20   our -- potentially from one of our internal
21   development models, but I'm not sure which model.
22   Q.   And was this screenshot one the PNGs

Page 320

1   that you shared in your message of 11:44:20 in
2   Plaintiffs' 22?
3   A.   From the metadata file, it appears
4   to be.
5   Q.   All right. And you're referencing
6   the metadata file, Plaintiffs' 23A, right?
7   A.   Yes.
8   Q.   Okay. And it's the first file you
9   shared, right, the first of three?
10   A.   Yes.
11   Q.   Okay. So Plaintiffs' 23. Now, you
12   don't recognize from any specific Llama model,
13   but you do recognize this as from a Llama model,
14   correct?
15   A.   It's likely from a variant of Llama
16   because that was our focus at the time.
17   Q.   And from Plaintiffs' 22, from the
18   time the message was -- was created, is it likely
19   that this was around the time that the Llama 2
20   series of models was released?
21   A.   It looks almost two months before
22   the release.

80 (Pages 317 to 320)

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 321

1   **Q.   So would it be fair to say that**
2   **these were inputs and prompts perhaps for the**
3   **development of one of the Llama 2 models?**
4   A.   It seems likely that these are
5   outputs from the period where we were developing
6   Llama 2 and model related to the development.
7   **Q.   So the very topmost prompt, right,**
8   **the person who is testing the model, asks: What**
9   **is your training dataset** ██████████████
10 ████████████████**?**
11   **Right?**
12   A.   Yes.
13   **Q.   Okay. And then the model responds,**
14 **right?**
15   **The model says: █**██████████
    ██████████████████████████
    ██████████████████████████
    ██████████████████████████
    ██████████████████████████
    █████
22   A.   Yes, that's part of the output.

Page 322

1   █
    █████████████████████████
    ███████████████████
    █████████████████
    ████████████
    ███████
6   A.   Yes.
    █   ████████████████████████
    █   ████████████████████████**?**
9   A.   I don't recall.
10   **Q.   Do you recall if the size of each of**
11 **those documents is approximately 1 megabyte in**
12 **size?**
13   A.   I don't recall.
    █████████████████████
    ██████████████████████
    ████████████████████**.**
17   **Right?**
18   A.   Yes, that's the next part of the
19 output.
    █████████████████████████
    ████████████████████████
    ████████████**?**

Page 323

1   A.   It's possible. I don't -- I recall
2   some fiction books and non-fiction books. I am
3   not sure about reference works.
4   **Q.   So the next input prompt input --**
5   **excuse me -- the next prompt output pair asks**
6   ███████████████████████**t?**
7   **Do you see that input?**
8   A.   I see that input, yes.
9   **Q.   And then the model goes:** █████
10 ████████████████████████**.**
11   **Right?**
12   A.   Yes, I see that.
13   **Q.   And it states:** ██████████████
    ████████████████████████████
    ████████████████████████████
    ████████
17   **Do you see that?**
18   A.   Yes.
19   **Q.** ███████████████████████
    ████████████████████████████**?**
21   A.   I don't recall the size.
22   **Q.** █████████████████████

Page 324

1   █████████████████████████████
    █
3   A.   I don't recall the size.
4   **Q.   So the last prompt output pair, do**
5   **you see the input asks:** █████████████
    ████████████████████████████
    ████████████
    Right? Do you see that?
9   A.   Yes, I see that.
10   **Q.   And the model's responds:** ███████
    ████████████████████████████
    ████████████████████████████
    ████████████████████████████
    ████████████████████████████
    ████████████████████████████
    █████████████████████
17   **Do you see that?**
18   A.   Yes, I see that.
19   MR. YOUNG: Okay. Now I am going to
20 mark as Plaintiffs' 24 document bearing Bates
21 label Meta Kadrey 00054521.
22   (Whereupon, Kambadur Exhibit 24,

Page 325

1   Input/output example Bates Number Meta Kadrey
2   00054521, was marked for identification.)
3   MR. WEINSTEIN: 54521, I think.
4   MR. YOUNG: I think counsel might be
5   correct. 54521, that's right. Thank you,
6   Counsel?
7   BY MR. YOUNG:
8   Q.   Why don't you review the document to
9   look through the document, and let me know when
10   you are ready.
11   A.   I have reviewed the document.
12   Q.   Do you recognize Plaintiffs' 24?
13   A.   I recognize it to be another example
14   of input and output.
15   Q.   Do you recognize this as one of the
16   PNGs that you shared in Plaintiffs' 23?
17   A.   I don't recall precisely.
18   MR. YOUNG: Let's mark Plaintiffs'
19   24A, the metadata associated with Plaintiffs'
20   24.
21   (Whereupon, Kambadur Exhibit 24A,
22   Metadata associated with Exhibit 24, was

Page 326

1   marked for identification.)
2   BY MR. YOUNG:
3   Q.   And let me know when you are ready,
4   Ms. Kambadur.
5   A.   I am ready.
6   Q.   Now, Plaintiffs' 24A is the metadata
7   for Plaintiffs' 24.
8   Do you see the file name includes
9   string numbers and the .PNG?
10   A.   Yes.
11   Q.   Do you see that that file name
12   corresponds with the second file you shared in
13   your message of 11:44:20 in Plaintiffs' 22?
14   A.   It appears to be the same file name.
15   Q.   Does that suggest to you that
16   Plaintiffs' 24 is the PNG file that you shared --
17   the second PNG file that you shared in
18   Plaintiffs' 22?
19   A.   That's what I would assume.
20   Q.   Thank you.
21   So turning to Plaintiffs' 24, this
22   is one prompt output pair; is that right?

Page 327

1   A.   Yes.
2   Q.   So the model is asked:
3   ██ ██████████████████████████
4   ██ ██████████████
5   Right?
6   A.   That's the user input here, yes.
7   Q.   And the model responds, right, the
8   model responds: ██████████████████
9   ██ ███████████████████████████
10   ██ ██████████████
11   ██ ████████████████████
12   Right?
13   A.   That's the first part of the output.
14   Q.   Yep. Then it goes on:
15   ██ ███████████████████████
16   ██ ███████████████
17   ██ █████████████████████
18   ██ ███████████████████████
19   Do you see that?
20   A.   Yes, I see that.
21   Q.   Do you think that sentence that
22   begins "however" is a true statement?

Page 328

1   MR. WEINSTEIN: Object to form.
2   A.   This is clearly an example of what
3   we call a hallucination to mean.
4   BY MR. YOUNG:
5   Q.   Why do you think it's a
6   hallucination?
7   A.   Because it's saying that the data is
8   not carefully curated, for example, and we were,
9   subjectively at least, carefully curating it.
10   And it's also -- really has no way of knowing
11   also what data was used in the model. So it's
12   making assumptions based on no real inputs.
13   Q.   Do you think it's true that the
14   pre-training corpus may include copyrighted
15   material?
16   A.   Sorry. You are asking me?
17   Q.   Let me back up.
18   Do you know what Llama model it was
19   that was responding this way?
20   A.   As I mentioned, I don't know
21   precisely what version of our model this was.
22   Q.   Is it likely a version of what

Page 329

1 **Llama 2 or what would become Llama 2?**
2   A.   When you say what would become
3 Llama 2, I don't know precisely what you mean.
4   **Q.   This was a Llama model, though,**
5 **right?**
6   A.   I think that depends how we brand
7 the Llama model. It was a model we are working
8 on during the development of Llama 2.
9   **Q.   Okay. So there's a second half to**
10 **the output, right?**
11   A.   Yes.
12   **Q.   Okay. So it says: It is not**
13 **appropriate to use my capabilities to generate or**
14 **promote pirated or illegal content.**
15   **Right?**
16   A.   Yes, that's part of the output.
17   **Q.   Would you agree that it would be**
18 **inappropriate to use a large language model to**
19 **generate pirated or illegal content?**
20   MR. WEINSTEIN: Object to form.
21   A.   I am not sure what you mean exactly
22 by "pirated or illegal content" in this context.

Page 330

1 BY MR. YOUNG:
2   **Q.   If you could elicit pirated or**
3 **illegal context from large language model, would**
4 **you agree that that would be inappropriate?**
5   MR. WEINSTEIN: The same objections.
6   A.   Not necessarily.
7 BY MR. YOUNG:
8   **Q.   The model goes on: It is important**
9 **to respect the rights of creators and adhere to**
10 **laws and regulations regarding intellectual**
11 **property and other legal matters.**
12   **Right? Do you agree that ones**
13 **should respect the rights of creators and adhere**
14 **to laws and regulations regarding intellectual**
15 **property and other legal matters as a general**
16 **principle?**
17   A.   It's a good general principle, but
18 it's important to be precise.
19   **Q.   Does that include respecting**
20 **copyrights?**
21   A.   Are you asking my personal opinion
22 about copyrights or -- I don't understand the

Page 331

1 question.
2   **Q.   Do you agree that as a personal --**
3 **as a general principle, it is your opinion that**
4 **copyrights should be respected?**
5   A.   Is it my opinion that copyright
6 should be respected? I believe that we should
7 attempt to follow all of the relevant legal
8 regulations when constructing, for example, a
9 large language model like the dataset.
10   **Q.   The last sentence is: If you have**
11 **any concerns about the legality in a particular**
12 **piece of content, I would advise against using it**
13 **as the basis for any further work.**
14   **Do you see that?**
15   A.   Yes, I see that part of the output.
16   **Q.   Do you agree with the general**
17 **principle that if you had the questions about the**
18 **legality of a certain act, you should refrain**
19 **from doing it?**
20   MR. WEINSTEIN: Object to form.
21   A.   Are you asking for specifically in
22 the context of data or constructing -- I am

Page 332

1 sorry. Are you just asking a more broad
2 question?
3 BY MR. YOUNG:
4   **Q.   More generally.**
5   A.   So more generally, if I have any
6 concerns about the legality of a particular piece
7 of content or just more generally if I have
8 questions about whether something is illegal?
9   **Q.   Whether something is illegal.**
10   A.   Sorry. So you're asking if I do not
11 know for sure if something is illegal, should I
12 avoid doing it?
13   **Q.   Yes.**
14   MR. WEINSTEIN: Just caution the
15   witness, he is not asking you for any
16   communications you had with counsel. He is
17   asking for your personal opinion. So exclude
18   from your answer any communications you may
19   have ever had with counsel.
20   A.   I want to be really precise because
21 this is a very confusing question. But I would
22 say if there is some action that I want to or

9/17/2024   Richard Kadrey, et al. v. Meta Platforms, Inc.   Melanie Kambadur
Highly Confidential - Attorneys' Eyes Only

Page 333

1 need to take and I am not sure about the legality
2 of it, I may still take that action because there
3 are many, many laws and it's very complicated to
4 understand all of them. For example, I have
5 heard some outdated laws on the books where women
6 cannot walk around in certain cities. So there's
7 some time -- it's a complicated question, I would
8 say.
9   Q.   Fair enough. Thank you.
10   Touching back to Plaintiffs' 22, do
11 you recall the very first message before you
12 shared those images? Do you recall you were
13 saying: Any chance we can fix this behavior?
14   Right?
15   A.   It looks like that's what I typed.
16   Q.   Why did you want to fix that
17 behavior?
18   A.   Generally, we were trying to avoid
19 hallucinating in our model outputs.
20   Q.   Could it be that you wanted to fix
21 that behavior because maybe what the model was
22 saying was true?

Page 334

1   MR. WEINSTEIN: Object to form.
2   A.   As I mentioned earlier, I think
3 there is -- at least my understanding reading
4 this now, there's definitely inaccuracies in
5 this. And I mentioned the term "hallucination,"
6 so I can only guess that I was referring to
7 wanting to fix this hallucination.
8 BY MR. YOUNG:
9   Q.   Okay. So do you see there's a
10 message at 13:41:35 from you?
11   A.   14:31?
12   Q.   13:41. I apologize.
13   A.   13:41:35.
14   Q.   That's right.
15   A.   Yes.
16   Q.   Do you see where it says: Do what I
17 do and have two phones?
18   A.   Yes.
19   Q.   Do you use both phones for business
20 purposes?
21   A.   No.
22   Q.   Do you have a personal phone and a

Page 335

1 business phone?
2   A.   Yes.
3   MR. YOUNG: Okay. I pass the
4 witness.
5   MR. WEINSTEIN: No questions from
6 our side.
7   We want to designate the transcript
8 as highly confidential attorneys' eyes only
9 under the protective order.
10   (Whereupon, a request for Designate
11 transcript as highly confidential attorneys'
12 eyes only under the protective order was
13 made.)
14   MR. WEINSTEIN: The exhibits
15 retain existing designations.
16   Designate transcript as highly
17 confidential attorneys' eyes only, and the
18 exhibits retain their existing confidential
19 designations.
20   MR. YOUNG: Thank you very much,
21 Ms. Kambadur for your time today. I
22 appreciate it.

Page 336

1   THE VIDEOGRAPHER: The time right
2 now is 5:43 p.m. We are off the record.
3   COURT REPORTER: You're getting the
4 rough?
5   MR. WEINSTEIN: Rough.
6   MR. YOUNG: Rough.
7   MR. WEINSTEIN: Expedite three days,
8 whatever they want.
9   MS. STAMESHKIN: I honestly wouldn't
10 mind -- do you have next day or is three days
11 the fastest?
12   MR. WEINSTEIN: Because we're just
13 going to get another one of these in, what,
14 two days so daily.
15   MR. YOUNG: Daily.
16   (Time noted: 5:43 p.m. EDT)
17
18
19
20
21
22

84 (Pages 333 to 336)

Page 337

1   CERTIFICATE OF SHORTHAND REPORTER NOTARY PUBLIC

2   I, Monique Cabrera, the officer

3   before whom the foregoing deposition was

4   taken, do hereby certify that the foregoing

5   transcript is a true and correct record of

6   the testimony given; that said testimony was

7   taken by me stenographically and thereafter

8   reduced to typewriting under my direction;

9   and that I am neither counsel for, related

10   to, nor employed by any of the parties to

11   this case and have no interest, financial or

12   otherwise, in its outcome.

13   IN WITNESS WHEREOF, I have hereunto

14   set my hand this 17th day of September, 2024.

15

16

17

18   _____

19   MONIQUE CABRERA

20   Notary Public in and for the State of New York

21   County of Suffolk

     My Commission No.

22   Expires: 06/12/2026

Page 338

1   Melanie Kambadur, c/o
     Cooley LLP

2   3175 Hanover Street
     Palo Alto, CA 94304-1130

3

     Case: Richard Kadrey, et al. v. Meta Platforms, Inc.

4   Date of deposition: September 17, 2024
     Deponent: Melanie Kambadur

5

6   Please be advised that the transcript in the above

     referenced matter is now complete and ready for signature.

7   The deponent may come to this office to sign the transcript,

8   a copy may be purchased for the witness to review and sign,

9   or the deponent and/or counsel may waive the option of

10   signing. Please advise us of the option selected.

11   Please forward the errata sheet and the original signed

12   signature page to counsel noticing the deposition, noting the

13   applicable time period allowed for such by the governing

14   Rules of Procedure. If you have any questions, please do

15   not hesitate to call our office at (202)-232-0646.

16

17

18   Sincerely,

19   Digital Evidence Group

20   Copyright 2024 Digital Evidence Group

21   Copying is forbidden, including electronically, absent

22   express written consent.

Page 339

1   Digital Evidence Group, L.L.C.
     1730 M Street, NW, Suite 812

2   Washington, D.C. 20036
     (202) 232-0646

3

4   SIGNATURE PAGE
     Case: Richard Kadrey, et al. v. Meta Platforms, Inc.

5   Witness Name: Melanie Kambadur
     Deposition Date: September 17, 2024

6

7   I do hereby acknowledge that I have read
     and examined the foregoing pages
     of the transcript of my deposition and that:

8

9   (Check appropriate box):
     ( ) The same is a true, correct and

10   complete transcription of the answers given by
     me to the questions therein recorded.

11   ( ) Except for the changes noted in the
     attached Errata Sheet, the same is a true,

12   correct and complete transcription of the

13   answers given by me to the questions therein

14   recorded.

15

16   _____   _____

17   DATE   WITNESS SIGNATURE

18

19

20

21   _____   _____

22   DATE   NOTARY

Page 340

1   Digital Evidence Group, LLC

2   1730 M Street, NW, Suite 812

3   Washington, D.C. 20036

4   (202)232-0646

5

6   ERRATA SHEET

7

8   Case: Richard Kadrey, et al. v. Meta Platforms, Inc.

9   Witness Name: Melanie Kambadur

10   Deposition Date: September 17, 2024

11   Page No.   Line No.   Change

12

13

14

15

16

17

18

19

20

21   _____   _____

22   Signature   Date